

APPROXIMATE ENERGY CONSERVATION OF SYMPLECTIC A-STABLE RUNGE–KUTTA METHODS FOR HAMILTONIAN SEMILINEAR EVOLUTION EQUATIONS

CLAUDIA WULFF AND MARCEL OLIVER

ABSTRACT. We prove that a class of A-stable symplectic Runge–Kutta time semidiscretizations (including the Gauss–Legendre methods) applied to a class of semilinear Hamiltonian PDEs which are well-posed on spaces of analytic functions with analytic initial data conserve a modified energy up to an exponentially small error. This modified energy is $O(h^p)$ -close to the original energy where p is the order of the method and h the time step-size. Examples of such systems are the semilinear wave equation or the nonlinear Schrödinger equation with analytic nonlinearity and periodic boundary conditions. Standard backward error analysis does not apply here because of the occurrence of unbounded operators in the construction of the modified vector field. This loss of regularity in the construction can be taken care of by projecting the PDE to a subspace where the operators occurring in the evolution equation are bounded and by coupling the number of excited modes as well as the number of terms in the expansion of the vectorfield with the stepsize. This way we obtain exponential estimates of the form $O(\exp(-\beta/h^{1/(1+q)}))$ with $\beta > 0$ and $q \geq 0$; for the semilinear wave equation, $q = 1$, and for the nonlinear Schrödinger equation, $q = 2$.

CONTENTS

1. Introduction	2
2. Backward error analysis for Hamiltonian ODEs	5
2.1. Implicit Runge–Kutta methods for ODEs	5
2.2. Embedding of the method into a flow	6
2.3. Modified Hamiltonian	7
3. Semilinear Hamiltonian PDEs	9
3.1. Semilinear evolution equations	9
3.2. Differentiability of the semiflow	10
3.3. Hamiltonian structures on Hilbert spaces	11
3.4. Spaces of analytic functions	14
3.5. Functional setting for the semilinear wave equation	15
3.6. Functional setting for the nonlinear Schrödinger equation	16
3.7. Nonlocal Schrödinger equation on a star-shaped domain	17
4. A-stable Runge–Kutta methods on Hilbert spaces	17
5. Backward error analysis for time-semidiscretizations	20
5.1. Statement of the main result	20
5.2. Galerkin truncation	22
5.3. Embedding of semidiscretizations into a flow	24

Date: August 23, 2012.

5.4. Backward error analysis on Hilbert spaces	28
Acknowledgments	33
References	33

1. INTRODUCTION

When discretizing conservative dynamical systems, it is often desirable to preserve their invariants or proper discrete analogs of such invariants. It is usually not possible to find a method that preserves all invariants, but there are often methods which preserve invariants approximately over very long times.

Symplectic integrators are a class of such methods that has been extensively used and studied in the context of Hamiltonian ordinary differential equations (ODEs). They are designed to preserve the symplectic form of the Hamiltonian system exactly and it can be shown that they also approximately preserve energy over exponentially long times, see e.g. [14, 15, 19, 33] and Section 2 below. The main technical tool used in the proof of such results is backward error analysis. Backward error analysis seeks a *modified Hamiltonian vectorfield* which at time h (the stepsize of the numerical method) approximates the one-step integrator ψ^h with stepsize h to very high accuracy and, on the other hand, exactly preserves a modified Hamiltonian \tilde{H} which, for a method of order p , is $O(h^p)$ -close to the Hamiltonian H of the Hamiltonian system to be discretized.

Results of this type go back to Neishtadt [25], who showed that an analytic diffeomorphism which is $O(h)$ -close to identity can be embedded into a flow up to an exponentially small error $O(e^{-h_0/h})$ by taking an optimal choice of $N = O(1/h)$ averaging steps of a normal form transformation. Benettin and Giorgilli [1] give an alternative construction, more suited towards the application to numerical integrators, by matching a Taylor expansion of a symplectic diffeomorphism ψ^h with the formal power series expansion of the flow of an arbitrary h -dependent Hamiltonian vector field, truncating at some order N , carefully estimating the remainders, and concluding that the truncation is optimal when $N = O(1/h)$. This method has been used to prove approximate energy conservation of many classes of symplectic numerical methods [14, 15, 19, 33]. In particular symplectic Runge–Kutta discretizations $\psi^h(\cdot)$ of analytic Hamiltonian ODEs approximately conserve a modified energy $\tilde{H}(x)$ with exponentially small error $O(e^{-h_0/2h})$ over exponentially long times $O(e^{h_0/2h})$ provided that the discrete trajectory remains in a bounded set, see Section 2 and in particular Theorem 2.3 for a more detailed review of those results.

In the case of partial differential evolution equations (PDEs), the phase space is typically an infinite dimensional Hilbert space and the vector field contains unbounded operators, usually in the form of spatial derivatives. This poses serious difficulties for backward error analysis. The flow map ϕ^t of the PDE usually ceases to be analytic, or even differentiable with respect to h and typically the same applies to the numerical time- h map ψ^h used to semidiscretize the PDE in time; the best one can generally guarantee is continuity in h . In other words, the formal power series expansion of ψ^h in h now contains powers of unbounded operators, and this propagates into the formal series expansion of the modified vector field.

Note, that the analytical difficulties persist when analyzing full space-time discretizations of the problem. In this case, the unbounded operators corresponding become matrices with norms diverging to infinity as the spatial resolution increases. Consequently, the constant h_0 in the exponential error estimate $O(e^{-h_0/h})$ tends to 0 with increasing spatial resolution, so that backward error analysis for general initial data without the requirement of high regularity leads to severe restrictions in the time stepsize. In the case of hyperbolic problems such as the semilinear wave equations, the standard approach fails for step size ratios close to the CFL limit, i.e., in the practically relevant regime.

Nonetheless, we can show that backward error analysis of this type can be used for PDEs provided the solution to the full PDE has high regularity. Time semidiscretizations of semilinear Hamiltonian PDEs are typically only well-defined if the method is fully implicit; explicit or partially implicit methods such as partitioned Runge–Kutta methods, the simplest of which are the leap frog and symplectic Euler schemes, violate the CFL condition [17]. Thus, in this paper we consider a class of (implicit) symplectic A-stable Runge–Kutta methods which includes the Gauss–Legendre Runge–Kutta methods. The simplest of these methods is the implicit midpoint rule.

Our approach applies to a large class of semilinear Hamiltonian PDEs with analytic nonlinearities, including the semilinear wave equation and the nonlinear Schrödinger equation on the circle. Our main result, Theorem 5.1, can be paraphrased as follows. If a semilinear Hamiltonian evolution equation with energy $H(\cdot)$ is discretized by a symplectic A-stable Runge–Kutta method Ψ^h of order p , then there exists a modified energy $\tilde{H}(\cdot)$ which is defined for Gevrey regular data, is $O(h^p)$ close to $H(\cdot)$, and is conserved by the symplectic integrator Ψ^h with exponentially small error $O(\exp(-\beta/h^{1/(q+1)}))$. Here, $q > 0$ is determined by the PDE, while β depends on the PDE, the regularity of the initial data, and the integrator. This result implies that if the solution to the continuum problem remains Gevrey regular over a finite interval of time, then, by consistency of the scheme, the discretization in time preserves a modified energy with an exponentially small error on the same time interval (Corollary 5.5). Moreover, provided the numerical trajectory $U^n = (\Psi^h)^n(U^0)$ remains Gevrey regular, the numerical energy is approximately conserved over exponentially long times (Remark 5.2).

Let us mention some related results: Moore and Reich [24] made a first step towards a formal backward error analysis for multisymplectic discretizations of semilinear wave equations and their corresponding energy and momentum conservation laws. They derive a modified higher order multisymplectic partial differential equation which is satisfied by the numerical solution with higher accuracy than the discretization error. See also the related work of Islas and Schober [18]. In [26], it is shown that second order finite difference space-semidiscretizations of analytic solutions of the semilinear wave equation approximately conserve a discrete momentum map up to an exponentially small error. Cano [3] considered symplectic space-time discretizations of semilinear wave equations and constructed a finite order modified Hamiltonian, assuming certain conjectures on the smoothness of the fully discrete system.

There has been a lot of recent progress on the approximate conservation of invariants by geometric integrators for Hamiltonian PDEs based on normal form transformations or modulated Fourier expansions. These results are strong in the

sense that they provide a detailed analysis which does not depend on extraneous regularity assumptions, however, by the techniques of proof are limited to small initial data, i.e., they apply in a weakly nonlinear regime, or they require a modified modified splitting method which dampens highly oscillatory motion.

Dujardin and Faou [6] prove approximate energy conservation of certain splitting methods applied to linear Schrödinger equations. Due to the linearity of the PDE, regularity of the numerical solution, if satisfied at the initial time, is guaranteed for all times. As a consequence, the numerical method ψ^h conserves a modified Hamiltonian \tilde{H} over exponentially long times. Debusche and Faou [5] prove that splitting methods applied to linear Schrödinger equations, where the Laplacian part is discretized by the implicit midpoint rule, conserve a modified energy exactly. Note that, in contrast, (symplectic) Gauss–Legendre Runge–Kutta discretizations of linear Hamiltonian systems preserve energy exactly [15].

In the nonlinear case, Faou, Grébert, and Paturel [8] prove conservation of actions of numerical trajectories of Hamiltonian evolution equations associated with splitting methods for small initial data over polynomially long times under generic non-resonance conditions for full-space time discretizations by using normal form theory under a condition coupling spatial and time stepsize that excludes time-semidiscretizations. For “rounded” splitting method which involve a cut-off in the high frequencies, they prove a similar result for time-semidiscretizations which implies long-time existence of regular numerical trajectories [9]. Faou and Grébert [7] show approximate energy conservation over exponentially long times for new splitting methods which dampen highly oscillatory terms of the linear part of the Hamiltonian evolution equation.

Cohen *et al.* [4] use modulated Fourier expansions to prove approximate conservation of energy and of the actions for semilinear wave equations over polynomially long times $O(1/h^k)$ with small smooth initial data and full space-time discretizations which are symplectic trigonometric integrators in time and spectral space discretizations. They do not obtain exponentially accurate estimates; but, assuming a non-resonance conditions which we do not require, they obtain boundedness of the numerical solution over polynomially long times, using the method of modulated Fourier expansions. A similar result for nonlinear Schrödinger equations is obtained by Gauckler and Lubich [12].

In the case of nonlinear ordinary differential equations, a numerical method ψ^h conserves a modified Hamiltonian in general only up to an exponentially small error [15]; error bounds over exponentially long times require the assumption of boundedness of the numerical solution in a suitable norm which is in general not satisfied for nonlinear PDEs. So the analysis of approximate energy conservation of symplectic discretizations of PDEs entails two problems: one is the construction of a modified Hamiltonian which is conserved by the numerical method with an error which is of high order for sufficiently smooth initial data (in this paper we are only interested in exponentially small errors); the second problem is to ensure that there exist numerical trajectories which remain bounded in the required norms over exponentially long times (see also the discussion above). We decouple the question of long time existence of numerical trajectory and existence of modified Hamiltonian. Although this paper primarily addresses the latter question, we think it is conceivable to obtain Nekhoroshev type estimates which imply existence of Gevrey regular numerical trajectories of Runge Kutta semidiscretizations in time

of semilinear Schrödinger equations over exponential long times. Such results have been obtained for the continuous problem in [10]. It might also be possible to obtain existence results for Gevrey-regular finite-dimensional invariant tori of numerical methods the existence of which has been obtained by Kuksin and Pöschel [30] and Pöschel [31]. Combined with our results, this would imply approximate energy conservation of A-stable Runge-Kutta semidiscretizations in time over exponential long times.

The method we use is related to techniques for averaging of rapidly forced Hamiltonian PDEs of Matthies and Scheel [23]. The analysis relies on our earlier work [27, 28] on A-stable Runge-Kutta methods for semilinear evolution equations: In [27], we analyzed the differentiability properties in the initial value and in time of the semiflow of

$$\partial_t U = F(U) = AU + B(U) \quad (1.1)$$

on a scale of Banach spaces, and obtained analogous results for the time- h map of its corresponding A-stable Runge-Kutta time-semidiscretization. In [28], we prove stability of the semiflow and of the time-semidiscrete solution under spatial spectral Galerkin approximation.

The paper is structured as follows. In Section 2 we review backward error analysis for symplectic discretizations of Hamiltonian ODEs. In Section 3 we introduce in more detail the class of semilinear Hamiltonian PDEs for which we carry out our error analysis and show that the semilinear wave equation and the nonlinear Schrödinger equation belong to this class. In Section 4 we introduce A-stable symplectic Runge-Kutta methods which are well-defined on Hilbert spaces. In Section 5, we present and prove our main results, Theorem 5.1 on approximate energy conservation of symplectic A-stable Runge-Kutta methods, as described in Section 4, applied to the class of PDEs introduced in Section 3.

2. BACKWARD ERROR ANALYSIS FOR HAMILTONIAN ODES

In the following, we review the method of backward error analysis for symplectic Runge-Kutta methods applied to Hamiltonian ordinary differential equations. We largely follow the notation and presentation of [15], but make explicit the dependence of the estimates on the magnitude of the vector field.

Throughout the paper, we write

$$\mathcal{B}_R^\mathcal{X}(U^0) = \{U \in \mathcal{X} : \|U - U^0\|_\mathcal{X} \leq R\}$$

to denote the closed ball of radius R in a Banach space \mathcal{X} about $U^0 \in \mathcal{X}$. When no confusion about the space is possible, we may drop the superscript \mathcal{X} .

2.1. Implicit Runge-Kutta methods for ODEs. Consider an autonomous ordinary differential equation

$$\dot{y} = f(y) \quad (2.1)$$

defined on the closed ball $\mathcal{B}_r(y^0) \subset \mathbb{R}^m$ of radius r around y^0 where f is analytic and satisfies the estimate

$$\|f(y)\| \leq M \quad \text{for } y \in \mathcal{B}_r(y^0).$$

Let $\phi^t : \mathcal{B}_r(y^0) \rightarrow \mathbb{R}^m$ denote the flow of (2.1), i.e. $y(t) = \phi^t(y^0)$ solves (2.1) with $y(0) = y^0$, and let $\psi^h : \mathcal{B}_r(y^0) \rightarrow \mathbb{R}^m$ denote a one-step method applied to (2.1).

As shown in [15, Theorem IX.7.2] (with r replaced by $2r$) via Cauchy estimates, the s -stage Runge–Kutta method

$$w^i = y^0 + h \sum_{j=1}^s a_{ij} f(w^j) \quad \text{for } i = 1, \dots, s, \quad (2.2a)$$

$$\psi^h(y^0) = y^0 + \sum_{i=1}^s b_i f(w^i) \quad (2.2b)$$

can be expanded in a converging power series in h ,

$$\psi^h(y) = y + \sum_{j=1}^{\infty} h^j g^j(y). \quad (2.3)$$

For future reference, we define the method-dependent constants

$$\eta = 2 \max(\|a\|, \|b\|/(2 \ln 2 - 1)) \quad \text{and} \quad \gamma = e(2 + 1.65\eta + \|b\|)$$

where

$$\|b\| = \sum_{i=1}^s |b_i| \quad \text{and} \quad \|a\| = \max_{i=1, \dots, s} \sum_{j=1}^s |a_{ij}|$$

2.2. Embedding of the method into a flow. The key result on which backward error analysis of such one-step methods is based is the embedding of the numerical time- h map into the flow of a modified vector field. The precise statement is the following.

Theorem 2.1 ([15, Theorem IX.7.6]). *In the notation of Section 2.1, let ψ^h be an s -stage Runge–Kutta method applied to (2.1). Then for every $h \in [0, h_0/4]$ with $h_0 = r/(2e\eta M)$ there exists a modified differential equation $\dot{y} = \tilde{f}(y)$, defined on $\mathcal{B}_{r/4}(y^0)$ whose flow $\tilde{\phi}^t$ satisfies $\tilde{\phi}^t(y^0) \in \mathcal{B}_{r/4}(y^0)$ for at least $0 \leq t \leq h$ and*

$$\|\psi^h(y^0) - \tilde{\phi}^h(y^0)\| \leq h \gamma M e^{-h_0/h}.$$

The proof shall not be repeated in detail here. But we note, for later reference, that the modified vector field is constructed as a power series in h ,

$$\tilde{f}^n(y; h) = f(y) + \sum_{j=p}^{n-1} h^j f^{j+1}(y), \quad (2.4)$$

where p is the order of the numerical method. Its exponential map is then expanded in powers of h and matched term by term with the expansion of the numerical time- h map (2.3). This yields a recursive expression for the coefficient vector fields [15, Lemma IX.7.3]:

$$f^j(y) = g^j(y) - \sum_{i=2}^j \frac{1}{i!} \sum_{k_1 + \dots + k_i = j} (D_{k_1} \cdots D_{k_{i-1}} f^{k_i})(y), \quad (2.5)$$

for $j \geq 2$ where $k_i \geq 1$ for all i , and $D_i g(y) = Dg(y) f^i(y)$ is a short hand notation for the Lie derivative with respect to the i th coefficient vector field. The proof of Theorem 2.1 proceeds by carefully estimating the growth of the f^j , noting that the optimal truncation is achieved when $n = n(h) = \lfloor h_0/h \rfloor$. (For $r \in \mathbb{R}$, the number $n = \lfloor r \rfloor$ denotes the largest integer $n \leq r$.) When referring to the optimally truncated vector field, we write \tilde{f}_h or just \tilde{f} .

One crucial estimate, as proved in [15], reads

$$\|\tilde{f}(y)\| \leq (1 + 1.65\eta) M \quad \text{for } y \in \mathcal{B}_{r/4}(y^0). \quad (2.6)$$

A related estimate, which guarantees consistency of the truncation, is the following. We provide an explicit proof as the dependence of the estimate on M and r is not spelled out in the literature in the detail we require.

Lemma 2.2. *In the notation of Section 2.1, let $p \geq 1$ and $0 \leq h \leq h_0/4$. Then there exists a constant $c = c(\eta, p)$ which only depends on the method such that*

$$\|\tilde{f}(y) - f(y)\| \leq c r^{-p} M^{p+1} h^p \quad \text{for } y \in \mathcal{B}_{r/4}(y^0).$$

Proof. It is known [15, Theorem IX.7.5] that

$$\|f^j(y)\| \leq \ln 2 \eta M \left(\frac{2\eta M j}{r} \right)^{j-1} \quad \text{for } y \in \mathcal{B}_{r/4}(y^0). \quad (2.7)$$

Applying this estimate to (2.4) and using that $n \leq h_0/h$ and therefore $h \leq h_0/n$, we find that

$$\begin{aligned} \|\tilde{f}(y) - f(y)\| &\leq h^p \sum_{j=p}^{n-1} h^{j-p} \|f^{j+1}(y)\| \\ &\leq h^p \sum_{j=p}^{n-1} h^{j-p} \ln(2) \eta M \left(\frac{2\eta M(j+1)}{r} \right)^j \\ &\leq h^p \ln(2) \eta M \sum_{j=p}^{n-1} \left(\frac{r}{2e\eta M n} \right)^{j-p} \left(\frac{2\eta M(j+1)}{r} \right)^j \\ &= \ln(2) \eta M \left(\frac{2h\eta M}{r} \right)^p e^{p+1} \sum_{j=p}^{n-1} \left(\frac{j+1}{n} \right)^{j-p} \frac{(j+1)^p}{e^{j+1}} \\ &\leq \ln(2) \eta M \left(\frac{2\eta M h}{r} \right)^p e^{p+1} p!, \end{aligned} \quad (2.8)$$

where, in the last inequality, we have bounded the first factor inside the sum by 1 and noted that $j^p e^{-j}$ is decreasing for $j \geq p$, so that

$$\sum_{j=p}^{n-1} \frac{(j+1)^p}{e^{j+1}} \leq \int_p^n x^p e^{-x} dx \leq \int_0^\infty x^p e^{-x} dx = p!.$$

This completes the proof. \square

2.3. Modified Hamiltonian. To proceed further, we assume that the differential equation is Hamiltonian, i.e. of the form

$$\dot{y} = f_H(y) = \mathbb{J} \nabla H(y) \quad (2.9)$$

with m even, where \mathbb{J} is a skew-symmetric, invertible $m \times m$ matrix, the *symplectic structure matrix*. The function $H: \mathbb{R}^n \rightarrow \mathbb{R}$ is the *Hamiltonian* or energy of the system and is an invariant of the motion. Hamiltonian systems have a symplectic flow map, i.e., ϕ^t satisfies

$$(\mathrm{D}_y \phi^t(y))^T \mathbb{J}^{-1} \mathrm{D}_y \phi^t(y) = \mathbb{J}^{-1} \quad (2.10)$$

for all y and t for which this relation is well defined [21]. A numerical one-step method is called symplectic if, when applied to a Hamiltonian ODE, its time- h map ψ^h is symplectic.

It is known that a Runge–Kutta method of the form (2.2) is symplectic if its coefficients satisfy

$$\mathbf{b}_i \mathbf{a}_{ij} + \mathbf{b}_j \mathbf{a}_{ji} - \mathbf{b}_i \mathbf{b}_j = 0$$

for $i, j = 1, \dots, s$; see, for example, [15]. The simplest example of a symplectic Runge–Kutta method is the *implicit midpoint rule*, given by

$$y^1 = y^0 + h f\left(\frac{y^0 + y^1}{2}\right),$$

which, equivalently, can be written in the form of a general Runge–Kutta scheme (2.2) with $s = 1$, $\mathbf{a}_{11} = \frac{1}{2}$, and $\mathbf{b}_1 = 1$.

For symplectic methods it can then be shown that the modified vector field is also Hamiltonian [14, 15, 19, 33]. More specifically, $\mathbb{J}^{-1}Df^j(y)$ is a symmetric matrix for all $j \geq p$ and $y \in \mathcal{B}_{r/4}(y^0)$, so that we can obtain a Hamiltonian for the vectorfield f^j via

$$H^j(y) = \int_0^1 \langle y - y^0, \mathbb{J}^{-1}f^j(y^0 + t(y - y^0)) \rangle dt \quad (2.11)$$

for $j \geq p$ and $\|y - y^0\| \leq r/4$, i.e., $f^j = \mathbb{J}\nabla H^j$. Then the optimally truncated modified Hamiltonian reads

$$\tilde{H}(\cdot; h) = H + \sum_{j=p}^{n-1} h^j H^{j+1}, \quad (2.12)$$

so that

$$\tilde{f} = \mathbb{J}\nabla \tilde{H}.$$

(Again, we refer to the optimal truncation by dropping the superscript of a general modified Hamiltonian \tilde{H}^n .) This construction implies the $O(h^p)$ -closeness of the modified to the original Hamiltonian. On the other hand, the modified Hamiltonian is exactly conserved on the modified flow which is exponentially close to the discrete time- h map. This proves approximate energy conservation of symplectic numerical methods.

Theorem 2.3 ([15, Theorem IX.8.1]). *Consider a Hamiltonian system (2.9) on some open, simply connected set $D \subset \mathbb{R}^{2m}$ with analytic Hamiltonian $H: D \rightarrow \mathbb{R}$ and apply a symplectic Runge–Kutta method ψ^h with stepsize h . So long as the numerical solution $y^n = (\psi^h)^n(y^0)$ remains in some compact set $K \subset D$ then, for h_0 and \tilde{H} as above,*

$$\tilde{H}(y^n) = \tilde{H}(y^0) + O(e^{-h_0/2h})$$

over the exponentially long interval of time $nh \leq e^{h_0/2h}$.

Together, (2.12) and Theorem 2.3 imply that the numerical method possess an $O(h^p)$ -approximate energy invariant which is preserved over time scales which are exponentially long in $1/h$.

In this paper we prove an analogue of Theorem 2.3 for a general class of general semilinear Hamiltonian evolution equations, including nonlinear Schrödinger equations and semilinear wave equations on bounded star-shaped domains. The main problem we face is that the assumptions of Section 2.1 cannot be satisfied with a finite value of M . We will address this problem in Section 5.

3. SEMILINEAR HAMILTONIAN PDES

We now provide the framework necessary to extend the results of the preceding section to the case of semilinear Hamiltonian systems on Hilbert spaces. We begin by reviewing the general functional setting for semilinear evolution equations. In Section 3.2 we review differentiability in time of the semiflow. In Section 3.3 we restrict to the Hamiltonian case and review a well-known integrability lemma in our Hilbert space setting. Section 3.4 introduces Hilbert spaces of analytic functions and superposition operators on these spaces. Finally, in Sections 3.5 and 3.6, respectively, we show how our main examples, the nonlinear Schrödinger equation and the semilinear wave equation, fit into this framework.

3.1. Semilinear evolution equations. We initially consider an abstract semilinear evolution equation of the form (1.1),

$$\partial_t U = F(U) = AU + B(U),$$

which formally resembles (2.1) but will be thought of as being posed on an infinite-dimensional Hilbert space \mathcal{Y} . To make this distinction apparent, we use upper case letters for phase space variables in infinite dimensions and lower case letters for phase space variables in finite dimensions. We assume the following.

(A) A is a normal operator on a Hilbert space \mathcal{Y} which generates a \mathcal{C}^0 -semigroup e^{tA} .

(B0) $B: \mathcal{D} \rightarrow \mathcal{Y}$ is analytic on an open set $\mathcal{D} \subset \mathcal{Y}$;

Recall that an operator A is normal if it is closed and $AA^* = A^*A$. For a definition of strongly continuous semigroups (\mathcal{C}^0 -semigroups), see [29].

Condition (A) implies that there is a constant $\omega > 0$ such that $\|e^{tA}\| \leq e^{t\omega}$ for all $t \geq 0$. Then, after casting (1.1) in its *mild formulation*

$$U(t) = e^{tA}U^0 + \int_0^t e^{(t-s)A} B(U(s)) ds, \quad (3.1)$$

we can apply the contraction mapping theorem with parameters to obtain well-posedness locally in time [16, 29]. Let $U^0 \rightarrow \Phi^t(U^0)$ denote the flow of (3.1), i.e., $U(t) = \Phi^t(U^0)$ satisfies (3.1) with $U(0) = U^0$.

For $m \in \mathbb{N}$, let \mathbb{P}_m denote the sequence of spectral projectors of A onto the set $\text{spec } A \cap \mathcal{B}_m^{\mathcal{C}}(0)$, set $\mathbb{P} \equiv \mathbb{P}_1$, and $\mathbb{Q} \equiv 1 - \mathbb{P}$. Assumption (A) implies that

$$\lim_{m \rightarrow \infty} \mathbb{P}_m U = U$$

for all $U \in \mathcal{Y}$, and

$$\|A\mathbb{P}_m U\|_{\mathcal{Y}} \leq m \|\mathbb{P}_m U\|_{\mathcal{Y}} \quad (3.2)$$

for $m \in \mathbb{N}$. Let $q > 0$, $\tau \geq 0$ and $\ell \in \mathbb{N}_0$. Since A is normal, $|\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q})$ is a well-defined, generally unbounded and densely defined operator on \mathcal{Y} . We may thus introduce the abstract *Gevrey space*

$$\mathcal{Y}_{\tau,\ell} = \mathcal{Y}_{\tau,\ell,q} = D(|\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q})) \quad (3.3)$$

equipped with the scalar product

$$\begin{aligned} \langle U_1, U_2 \rangle_{\mathcal{Y}_{\tau,\ell}} &= \langle \mathbb{P}U_1, \mathbb{P}U_2 \rangle_{\mathcal{Y}} \\ &+ \langle |\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q}) \mathbb{Q}U_1, |\mathbb{Q}A|^\ell \exp(\tau|\mathbb{Q}A|^{1/q}) \mathbb{Q}U_2 \rangle_{\mathcal{Y}}. \end{aligned} \quad (3.4)$$

Gevrey-smooth functions $U \in \mathcal{Y}_{\tau,\ell}$ are exponentially well approximated by their Galerkin projections $\mathbb{P}_m U$. Indeed, setting $\mathbb{Q}_m = \text{id} - \mathbb{P}_m$,

$$\|\mathbb{Q}_m U\|_{\mathcal{Y}} \leq m^{-\ell} \exp(-\tau m^{1/q}) \|U\|_{\mathcal{Y}_{\tau,\ell}}. \quad (3.5)$$

Moreover, this definition of the norm ensures that

$$\|A\|_{\mathcal{Y}_{\tau,\ell+1} \rightarrow \mathcal{Y}_{\tau,\ell}} \leq 1 \quad \text{and} \quad \|U\|_{\mathcal{Y}_{\tau,\ell}} \leq \|U\|_{\mathcal{Y}_{\tau,\ell+1}} \quad (3.6)$$

for all $U \in \mathcal{Y}_{\tau,\ell+1}$. For convenience, we define $\mathcal{Y}_\ell \equiv \mathcal{Y}_{0,\ell}$. We can then state the following lemma.

Lemma 3.1. *For $\sigma > \tau$ and $p \in \mathbb{N}_0$,*

$$\|A^p\|_{\mathcal{Y}_{\sigma,\ell} \rightarrow \mathcal{Y}_{\tau,\ell}} \leq \left(\frac{pq}{e(\sigma - \tau)} \right)^{pq}. \quad (3.7)$$

Proof. For $p = 0$, there is nothing to prove, hence let $p > 0$. For fixed $U \in \mathcal{Y}_{\sigma,\ell}$,

$$\|A^p U\|_{\mathcal{Y}_{\tau,\ell}}^2 = \|\mathbb{P}U\|_{\mathcal{Y}}^2 + \| |\mathbb{Q}A|^p e^{(\tau-\sigma)|\mathbb{Q}A|^{1/q}} |\mathbb{Q}A|^\ell e^{\sigma|\mathbb{Q}A|^{1/q}} \mathbb{Q}U \|_{\mathcal{Y}}^2. \quad (3.8)$$

The function $f(\lambda) = |\lambda|^p e^{(\tau-\sigma)|\lambda|^{1/q}}$ is non-negative and has a global maximum at $\lambda_* = (pq/(\sigma - \tau))^q$. Replacing the corresponding term in (3.8) by its maximum value, we obtain (3.7). \square

3.2. Differentiability of the semiflow. To obtain a flow of the evolution equation which has higher order time derivatives, as required in Section 5, we need more specific assumptions on the regularity of B on the scale of Hilbert spaces defined above.

We use the following convention to denote derivatives. Given any Hilbert space \mathcal{X} , open set $\mathcal{D} \subset \mathcal{X}$, and map $Z: \mathcal{D} \rightarrow \mathbb{R}$, we write $DZ(U)$ to denote the derivative of Z at $U \in \mathcal{D}$ as an element of \mathcal{X}^* , and by $\nabla Z(U)$ the canonical representation of $DZ(U)$ by an element of \mathcal{X} . In other words, $DZ(U)W = \langle \nabla Z(U), W \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product on \mathcal{X} .

For Banach spaces \mathcal{X} and \mathcal{Z} , and $j \in \mathbb{N}_0$, we write $\mathcal{E}^j(\mathcal{Y}, \mathcal{X})$ to denote the vector space of j -multilinear bounded mappings from \mathcal{Y} to \mathcal{X} ; we set $\mathcal{E}^j(\mathcal{X}) \equiv \mathcal{E}^j(\mathcal{X}, \mathcal{X})$. Moreover, when $\mathcal{U} \subset \mathcal{X}$ is open and $k \in \mathbb{N}$, we write $\mathcal{C}_b^k(\mathcal{U}, \mathcal{Z})$ to denote the set of k times continuously differentiable functions $F: \mathcal{U} \rightarrow \mathcal{Z}$ whose derivatives $D^i F$ are bounded as maps from \mathcal{U} to $\mathcal{E}^i(\mathcal{X}, \mathcal{Z})$ and extend to the boundary of \mathcal{U} .

Finally, for Banach spaces \mathcal{X} , \mathcal{Y} , and \mathcal{Z} , and open subsets $\mathcal{U} \subset \mathcal{X}$, $\mathcal{V} \subset \mathcal{Y}$, and $\mathcal{W} \subset \mathcal{Z}$, we write

$$F \in \mathcal{C}_b^{(\underline{m}, n)}(\mathcal{U} \times \mathcal{V}; \mathcal{W})$$

to denote a continuous, bounded function $F: \mathcal{U} \times \mathcal{V} \rightarrow \mathcal{W}$ whose partial Fréchet derivatives $D_X^i D_Y^j F(X, Y)$ exist, are bounded, and are such that the maps

$$(X, Y, X_1, \dots, X_i) \mapsto D_X^i D_Y^j F(X, Y)(X_1, \dots, X_i)$$

are continuous from $\mathcal{U} \times \mathcal{V} \times \mathcal{X}^i$ into $\mathcal{E}^j(\mathcal{Y}, \mathcal{Z})$ for $i = 0, \dots, m$ and $j = 0, \dots, n$, and extend continuously to the boundary.

Given $\delta > 0$ and a family of open sets $\mathcal{D}_\ell \subset \mathcal{Y}_\ell$ for $\ell = 0, \dots, L$ for $L \in \mathbb{N}$, we define

$$\mathcal{D}_\ell^{-\delta} \equiv \{U \in \mathcal{D}_\ell: \text{dist}_{\mathcal{Y}_\ell}(U, \partial \mathcal{D}_\ell) > \delta\}. \quad (3.9)$$

Then, by construction, $\mathcal{B}_\delta^{\mathcal{Y}_\ell}(U) \subset \mathcal{D}_\ell$ for all $U \in \mathcal{D}_\ell^{-\delta}$ and $\ell = 0, \dots, L$.

Let \mathcal{Y}_1 be a Banach space continuously embedded into the Banach space \mathcal{Y} . Then $\mathcal{D}_1 \subset \mathcal{Y}_1$ is called a δ_* -nested subset of $\mathcal{D} \subset \mathcal{Y}$ if $\mathcal{D}_1^{-\delta} \subset \mathcal{D}^{-\delta}$ for all $\delta \in [0, \delta_*]$. Furthermore we say that the family $\mathcal{D}_0, \dots, \mathcal{D}_L$ is δ_* -nested if $\mathcal{D}_\ell^{-\delta} \subset \mathcal{D}_{\ell-1}^{-\delta}$ for all $\delta \in [0, \delta_*]$ with $\delta_* > 0$ and $\ell = 1, \dots, L$. For example, the family $\mathcal{D}_k = \mathcal{B}_R^{\mathcal{Y}_k}(U^0)$ is δ_* -nested for every $\delta_* \in (0, R)$ and $U^0 \in \mathcal{Y}_L$. However, an arbitrary nested family $\mathcal{D}_\ell \subset \mathcal{Y}_\ell$ may not be δ_* -nested for any $\delta_* > 0$.

We now make the following assumption on the nonlinearity of our semilinear evolution equation.

- (B1) There exist $K \in \mathbb{N}_0$, $N \in \mathbb{N}$ with $N > K$, and a δ_* -nested sequence of \mathcal{Y}_k -bounded and open sets \mathcal{D}_k such that $B \in \mathcal{C}_b^{N-k}(\mathcal{D}_k, \mathcal{Y}_k)$ for $k = 0, \dots, K$.

We denote the bounds of the maps $B: \mathcal{D}_k \rightarrow \mathcal{Y}_k$ and their derivatives by constants M_k, M'_k , etc., for $k = 0, \dots, K$. In addition to the domains $\mathcal{D}_0, \dots, \mathcal{D}_K$ defined in (B1), we will sometimes refer to \mathcal{D}_{K+1} which may be any \mathcal{Y}_{K+1} -bounded, open, and δ_* -nested subset of \mathcal{D}_K , and define

$$R_{K+1} = \sup_{U \in \mathcal{D}_{K+1}} \|U\|_{\mathcal{Y}_{K+1}}. \quad (3.10)$$

We can then quote the following theorem on the uniform regularity of the flow [27, Theorem 2.6 and Remark 2.8].

Theorem 3.2 (Regularity of semiflow). *Assume (A) and (B1). Choose $\delta \in (0, \delta_*]$ small enough such that $\mathcal{D}_{K+1}^{-\delta} \neq \emptyset$. Then there exists $T_* > 0$ such that the semiflow $(U, t) \mapsto \Phi^t(U)$ of (1.1) satisfies*

$$\Phi \in \bigcap_{\substack{j+k \leq N \\ \ell \leq k \leq K+1}} \mathcal{C}_b^{(j, \ell)}(\mathcal{D}_{K+1}^{-\delta} \times (0, T_*); \mathcal{Y}_{k-\ell}). \quad (3.11)$$

Moreover, Φ maps $\mathcal{D}_{K+1}^{-\delta} \times [0, T_*]$ into \mathcal{D}_K . The bounds on Φ and T_* depend only on the bounds afforded by (B1), (3.10), ω , and on δ .

Remark 3.3. The precise form of assumption (B1) is motivated by the typical case where B is a superposition operator of a function $f: D \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$ and \mathcal{Y}_ℓ is related to the standard Sobolev space \mathcal{H}_ℓ . Then, if f is $(N+1)$ times continuously differentiable on some open set $D \subset \mathbb{R}^d$, it is N times differentiable as a map from the open set \mathcal{D} of \mathcal{H}_1 to \mathcal{H}_1 . Here $u \in \mathcal{D}$ ensures that $u(x) \in D$ pointwise. Moreover f is $(N-k)$ times differentiable from $\mathcal{D} \cap \mathcal{H}_k$ to \mathcal{H}_k —see [27, Theorem 2.12] and also [26, Remark 7.4].

For the results of Section 5, we need regularity of the flow on a space of Gevrey-regular functions as well. Hence, we assume the following.

- (B2) There exist $\tau > 0$, $q > 0$, $L \geq 0$, and an $\mathcal{Y}_{\tau, L}$ -bounded open set $\mathcal{D}_{\tau, L}$ which is a δ_* -nested subset of \mathcal{D}_{K+1} such that $B \in \mathcal{C}_b^1(\mathcal{D}_{\tau, L}, \mathcal{Y}_{\tau, L})$.

We note that $\mathcal{Y}_{\tau, L} \subset \mathcal{Y}_{K+1}$ due to Lemma 3.1. In the following, $\mathcal{D}_{\tau, L+1}$ refers to an arbitrary $\mathcal{Y}_{\tau, L}$ -bounded, open, and δ_* -nested subset of $\mathcal{D}_{\tau, L}$. We note that under assumption (B2), Theorem 3.2 applies with $\mathcal{Y}_{\tau, L}$ in place of \mathcal{Y} . For examples, see Sections 3.5 and 3.6.

3.3. Hamiltonian structures on Hilbert spaces. Our main result, the backward error analysis of Section 5, requires that (1.1) is Hamiltonian, i.e., there exists

a symplectic structure operator \mathbb{J} and a Hamiltonian $H: \mathcal{D} \rightarrow \mathbb{R}$ such that

$$\partial_t U = AU + B(U) = \mathbb{J}\nabla H(U). \quad (3.12)$$

In addition to (A) and (B0), we assume the following.

- (H0) The symplectic structure operator \mathbb{J} is a closed, skew-symmetric, densely defined, and bijective linear operator on \mathcal{Y} .
- (H1) A is skew-symmetric and $\mathbb{J}^{-1}A$ is bounded and self-adjoint on \mathcal{Y} .
- (H2) For every $U \in \mathcal{D}$, the operator $\mathbb{J}^{-1}DB(U)$ is self-adjoint on \mathcal{Y} .
- (H3) \mathcal{D} is bounded and star-shaped.

We first remark that, by the closed graph theorem, (H0) implies that \mathbb{J} is invertible with $\mathbb{J}^{-1} \in \mathcal{E}(\mathcal{Y})$. This implies, in particular, that $\mathbb{J}^{-1}DB(U)$ is a bounded operator on \mathcal{Y} for every $U \in \mathcal{D}$.

Second, recall that an operator A is skew if $A^* = -A$ and $D(A) = D(A^*)$. This implies that $\text{spec}(A) \subset i\mathbb{R}$ and that, by Stone's Theorem, A generates a unitary \mathcal{C}^0 -group on \mathcal{Y} ; see, e.g. [32]. If $A = A_s + A_b$ where A_s is skew and A_b is bounded, we can redefine B as $B + A_b$ and A as A_s to satisfy (H1). This situation is typical for semilinear wave equations, see Section 3.5.

Third, by conditions (A), (H0), and (H1),

$$\mathbb{J}^{-1}A = (\mathbb{J}^{-1}A)^* = A^*\mathbb{J}^{-1} = -A\mathbb{J}^{-1} = A\mathbb{J}^{-1}. \quad (3.13)$$

Hence, A and \mathbb{J}^{-1} commute, which also implies the following.

Lemma 3.4. *Assume (A), (H0), and (H1). Then $\mathbb{J}^{-1}\mathbb{P}_m = \mathbb{P}_m\mathbb{J}^{-1}$ for all $m \in \mathbb{N}_0$.*

Proof. By (3.13), A and \mathbb{J}^{-1} commute, and so do $F(A)$ and \mathbb{J}^{-1} for all analytic functions F . Approximating characteristic functions χ_Λ of measurable sets $\Lambda \subset \mathbb{C}$ by analytic functions, we see that $\chi_\Lambda(A)$ and \mathbb{J}^{-1} also commute [32]. With $\Lambda = \mathcal{B}_m^{\mathbb{C}}(0)$, this implies that $\chi_\Lambda(A) = \mathbb{P}_m$ commutes with \mathbb{J}^{-1} . \square

Finally, we recall that a subset \mathcal{S} of a linear space is star-shaped if there exists $U_* \in \mathcal{S}$ such that for every $W \in \mathcal{S}$, the line segment U_*W is contained in \mathcal{S} . We then say that \mathcal{S} is star-shaped with respect to U_* ; the set of all points with respect to which \mathcal{S} is star-shaped is denoted $\text{ck } \mathcal{S}$, the *convex kernel* of \mathcal{S} .

The existence of a Hamiltonian H is then guaranteed by the following integrability lemma.

Lemma 3.5. *Assume (A), (B0) and (H0–3). Then there exists an analytic Hamiltonian $H: \mathcal{D} \rightarrow \mathbb{R}$ for the evolution equation (1.1). Moreover, if (B1) holds true, then $H \in \mathcal{C}_b^N(\mathcal{D}; \mathbb{R})$.*

Proof. We seek a Hamiltonian of the form

$$H(U) = \frac{1}{2} \langle U, \mathbb{J}^{-1}AU \rangle + V(U). \quad (3.14)$$

Due to (A), (H0), and (H1), the quadratic part of the Hamiltonian is well-defined and possesses the properties claimed.

To proceed, we first suppose that \mathcal{D} is star-shaped and fix $U^0 \in \text{ck } \mathcal{D}$. We set

$$V(U) = \int_0^1 \langle \mathbb{J}^{-1}B(tU + (1-t)U^0), U - U^0 \rangle dt, \quad (3.15)$$

so that, for $W \in \mathcal{Y}$,

$$\begin{aligned} \langle \nabla V(U), W \rangle &= \int_0^1 \langle \mathbb{J}^{-1} B(tU + (1-t)U^0), W \rangle dt \\ &\quad + t \int_0^1 \langle \mathbb{J}^{-1} DB(tU + (1-t)U^0)W, U - U^0 \rangle dt \\ &= \int_0^t \frac{d}{dt} \langle t \mathbb{J}^{-1} B(tU + (1-t)U^0), W \rangle dt, \end{aligned}$$

where the last equality is due to the self-adjointness of $\mathbb{J}^{-1}DB(U)$. Then, by the fundamental theorem of calculus, $B(U) = \mathbb{J}\nabla V(U)$. Further, (3.15) shows that analyticity of B implies analyticity of V , and uniform bounds on B imply corresponding uniform bounds on V . \square

In the construction of the modified Hamiltonian in Section 5, we will need to shrink domains of definition. Therefore, we need conditions under which the domains $\mathcal{D}_\ell^{-\delta}$ remain star-shaped. We make use of the following characterization of a star-shaped subset \mathcal{S} of a linear space [34]. Let $\{C_\alpha : \alpha \in A\}$ denote the collection of all maximal convex subsets of \mathcal{S} , which is nonempty. Then

$$\text{ck } \mathcal{S} = \bigcap_{\alpha \in A} C_\alpha \quad \text{and} \quad \mathcal{S} = \bigcup_{\alpha \in A} C_\alpha.$$

We now prove a result which shows that so long as the convex kernel remains nonempty under domain shrinking, the reduced set remains star-shaped.

Theorem 3.6. *Let \mathcal{Y} be a Banach space and $\mathcal{D} \subset \mathcal{Y}$ star-shaped with $(\text{ck } \mathcal{D})^{-\delta}$ nonempty. Then $\mathcal{D}^{-\delta}$ is star-shaped.*

Proof. Let $\{C_\alpha : \alpha \in A\}$ denote the collection of all maximal convex subsets of \mathcal{D} . Then $\{C_\alpha^{-\delta} : \alpha \in A\}$ is a collection of convex subsets of $\mathcal{S}^{-\delta}$. We claim that it is the collection of all maximal convex subsets of $\mathcal{S}^{-\delta}$ up to boundary points. (It is generally not true that a maximal convex subset of an open set is open, hence this qualification.)

To prove this claim, let E denote an arbitrary maximal convex subset of $\mathcal{S}^{-\delta}$, and set

$$E^{+\delta} = \bigcup_{x \in E} \text{int } \mathcal{B}_\delta(x), \quad (3.16)$$

where $\text{int } \mathcal{D}$ denotes the interior of a set \mathcal{D} . Let $F \supset E^{+\delta}$ denote a maximal convex extension of $E^{+\delta}$ in \mathcal{S} . Clearly, $\text{int } E \subset F^{-\delta} \subset \mathcal{S}^{-\delta}$.

Vice versa, suppose $x \in F^{-\delta} \setminus \text{cl } E$. Since the convex hull of $\{x\} \cup E$ is contained in $\mathcal{S}^{-\delta}$, this contradicts the maximality of E up to boundary points. We conclude that $\text{int } E = F^{-\delta}$. I.e., $\{C_\alpha^{-\delta} : \alpha \in A\}$ is the collection of all convex subsets of $\mathcal{S}^{-\delta}$ which are maximal up to boundary points, and

$$\bigcap_{\alpha \in A} C_\alpha^{-\delta} \subset \text{ck}(\mathcal{S}^{-\delta}) \subset \bigcap_{\alpha \in A} \text{cl } C_\alpha^{-\delta}.$$

Since, by assumption, $(\text{ck } \mathcal{S})^{-\delta}$ is nonempty, there exists an $x \in \text{ck } \mathcal{S}$ satisfying $\text{dist}(x, \partial \text{ck } \mathcal{S}) > \delta$. Clearly, $x \in C_\alpha$ with $\text{dist}(x, \partial C_\alpha) > \delta$ for every $\alpha \in A$. Thus, $x \in C_\alpha^{-\delta}$ and we conclude that $\text{ck}(\mathcal{S}^{-\delta})$ is also nonempty. This implies that $\mathcal{S}^{-\delta}$ is star-shaped. \square

For bounds on the modified Hamiltonian in Section 5 we will need (H3) for $\delta \in (0, \delta_*]$ on at least two scale rungs, so that, for simplicity, we assume the following.

- (H4) Each \mathcal{D}_k for $k = 0, \dots, K$ is bounded and star-shaped with $(\text{ck } \mathcal{D}_k)^{-\delta_*}$ non-empty.

The following are examples of hierarchies of sets which satisfy (H4) in general Banach spaces.

Example 3.7. The domains \mathcal{D}_k are open \mathcal{Y}_k balls $\text{int}(\mathcal{B}_{R_k}^{\mathcal{Y}_k}(0))$ with radii $R_K \leq \dots \leq R_0$.

Example 3.8. Let $\mathcal{D} = \mathcal{D}_0$ be a bounded star-shaped domain, fix $K \in \mathbb{N}$, and let $R > 0$ sufficiently large and $\delta_* > 0$ sufficiently small such that $(\text{ck } \mathcal{D})^{-\delta_* K} \neq \emptyset$, $\mathcal{D} \subset \mathcal{B}_R^{\mathcal{Y}}(U^0)$ for some $U^0 \in (\text{ck } \mathcal{D})^{-\delta_* K} \cap \mathcal{Y}_K$ and $R > K\delta_* > 0$. Then, for $k = 0, \dots, K$, the domains $\mathcal{D}_k \equiv \mathcal{D}^{-\delta_* k} \cap \text{int}(\mathcal{B}_{R-\delta_* k}^{\mathcal{Y}_k}(U^0))$ are nested and star-shaped with $(\text{ck } \mathcal{D}_k)^{-\delta_*} \neq \emptyset$. By Theorem 3.6, the sets $\mathcal{D}_k^{-\delta}$ are star-shaped for all $\delta \in [0, \delta_*]$ and, by construction,

$$\mathcal{D}_{k+1}^{-\delta} \subset \mathcal{D}^{-\delta_*(k+1)} \cap \mathcal{B}_{R-\delta_*(k+1)}^{\mathcal{Y}_{k+1}}(U^0) \subset \mathcal{D}^{-\delta_*(k+1)} \cap \mathcal{B}_{R-\delta_*(k+1)}^{\mathcal{Y}_k}(U^0) \subset \mathcal{D}_k^{-\delta}.$$

Hence, $\mathcal{D}_0, \dots, \mathcal{D}_K$ are δ_* -nested and (H4) holds.

In the next section we introduce concrete function spaces and superposition operators on these spaces in order to verify that our main examples, the nonlinear Schrödinger equation and the semilinear wave equation, fit into our abstract framework.

3.4. Spaces of analytic functions. We denote the Fourier coefficients of a function $u \in \mathcal{L}^2(\mathbb{S}^1; \mathbb{C}^d)$ by \hat{u}_k , so that

$$u(x) = \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \hat{u}_k e^{ikx}. \quad (3.17)$$

Let $\mathcal{G}_{\tau, \ell} \equiv \mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C}^d)$ denote the Hilbert space of analytic functions $u \in \mathcal{L}^2(\mathbb{S}^1; \mathbb{C}^d)$ for which

$$\|u\|_{\mathcal{G}_{\tau, \ell}}^2 \equiv \langle u, u \rangle_{\mathcal{G}_{\tau, \ell}} < \infty,$$

where the inner product is given by

$$\langle u, v \rangle_{\mathcal{G}_{\tau, \ell}} = \langle \hat{u}_0 \hat{v}_0 \rangle_{\mathbb{C}^d} + \sum_{k \in \mathbb{Z}} k^{2\ell} e^{2\tau|k|} \langle \hat{u}_k \hat{v}_k \rangle_{\mathbb{C}^d}. \quad (3.18)$$

It can be shown that $\mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C}^d)$ contains all real analytic functions whose radius of analyticity is at least τ . In particular, functions in $\mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C}^d)$ can be differentiated infinitely often. This follows from Lemma 3.1 with $\mathcal{Y} = \mathcal{L}^2(\mathbb{S}^1; \mathbb{C}^d)$ and $A = i\partial_x$. We write $\mathcal{H}^\ell \equiv \mathcal{G}_{0, \ell}$ to denote the usual Sobolev space of functions whose weak derivatives up to order ℓ are square-integrable.

The additional index ℓ in $\mathcal{G}_{\tau, \ell}$ is important because of the following.

Lemma 3.9 ([11, Lemma 1]). *The space $\mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C})$ is a topological algebra for every $\tau \geq 0$ and $\ell > 1/2$. Specifically, there exists a constant $c = c(\ell)$ such that for every $u, v \in \mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C})$ the product $uv \in \mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C})$ with*

$$\|uv\|_{\mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C})} \leq c \|u\|_{\mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C})} \|v\|_{\mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C})}. \quad (3.19)$$

To treat general nonlinear potentials in (3.22), we need to consider superposition operators $f: \mathcal{G}_{\tau,\ell} \rightarrow \mathcal{G}_{\tau,\ell}$ of analytic functions. The following lemma is a minor adaptation of results proved in [11, 22].

Lemma 3.10. *If $f: \mathbb{C}^d \rightarrow \mathbb{C}^d$ is entire then f is also entire as a function from $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1; \mathbb{C}^d)$ to itself for every $\tau \geq 0$ and $\ell > 1/2$. If f is analytic on $\mathcal{B}_r(u_0) \subset \mathbb{C}^d$ where $u_0 \in \mathbb{C}^d$ then f is analytic and uniformly bounded from $\mathcal{B}_R(u_0) \subset \mathcal{G}_{\tau,\ell}(\mathbb{S}^1; \mathbb{C}^d)$ to $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1; \mathbb{C}^d)$ for any $R < r/c$, where $c = c(\ell)$ is the constant from Lemma 3.9.*

Proof. Let f be entire and let

$$f(z) = \sum_{n=0}^{\infty} a_n (z - u_0)^n \quad (3.20)$$

be the Taylor series of f around $u_0 \in \mathbb{C}$. Let $\phi: \mathbb{R} \rightarrow \mathbb{R}$ be its majorization

$$\phi(s) = \sum_{n=0}^{\infty} |a_n| s^n.$$

By applying the algebra inequality (3.19) to each term of the power series expansion (3.20) of $f(u)$, we see that the series converges for every $u \in \mathcal{G}_{\tau,\ell}$ provided $\tau \geq 0$ and $\ell > 1/2$, and that

$$\|f(u)\|_{\mathcal{G}_{\tau,\ell}} \leq c^{-1} \phi(c \|u - u_0\|_{\mathcal{G}_{\tau,\ell}}) + |a_0|(\sqrt{2\pi} - c^{-1}), \quad (3.21)$$

where c is as in Lemma 3.9, see [11]. In other words, f is entire on $\mathcal{G}_{\tau,\ell}(\mathbb{S}^1)$.

When f has only a finite radius of analyticity, we argue as follows. Assume that $|f(z)| \leq M$ on $\mathcal{B}_r^{\mathbb{C}}(u_0)$. Then, by Cauchy's estimate,

$$|a_n| \leq \frac{M}{r^n}.$$

Consequently, the majorant ϕ is bounded on any $\mathcal{B}_\rho^{\mathbb{C}}(0)$ with $\rho < r$ with uniform bound

$$\mu = \frac{M}{1 - \rho/r}.$$

Due to (3.21), the superposition operator f is then analytic and bounded by

$$M_{\text{spp}} = \mu/c + |a_0|(\sqrt{2\pi} - c^{-1})$$

as a map from a ball of radius $R = \rho/c$ around $u_0 \in \mathcal{G}_{\tau,\ell}$ into $\mathcal{G}_{\tau,\ell}$. \square

3.5. Functional setting for the semilinear wave equation. Consider the semilinear wave equation

$$\partial_{tt}u = \partial_{xx}u - V'(u) \quad (3.22)$$

on the circle $\mathbb{S}^1 = \mathbb{R}/2\pi\mathbb{Z}$. Its Hamiltonian can be written

$$H(u, v) = \int_{\mathbb{S}^1} \left[\frac{1}{2} v^2 + \frac{1}{2} (\partial_x u)^2 + V(u) \right] dx$$

where $v = \partial_t u$. We write $U = (u, v)^T$ and set $\mathcal{Y} \equiv \mathcal{H}_1(\mathbb{S}^1; \mathbb{R}) \times \mathcal{L}^2(\mathbb{S}^1; \mathbb{R})$ so that the Hamiltonian is well-defined on \mathcal{Y} . For $U = (u, v) \in \mathcal{Y}$ let $\mathbb{P}_0 U = (p_0 u, p_0 v)$ where for $u \in \mathcal{L}^2(\mathbb{S}^1; \mathbb{R})$ we define $p_0 u = \hat{u}_0$. Setting

$$\tilde{A} = \begin{pmatrix} 0 & \text{id} \\ \partial_x^2 & 0 \end{pmatrix},$$

we then define

$$A = \mathbb{P}_0 \tilde{A}, \quad B(U) = \begin{pmatrix} 0 \\ -V'(u) \end{pmatrix} + \mathbb{P}_0 \tilde{A}U, \quad (3.23)$$

and the symplectic structure matrix \mathbb{J} via

$$\langle \mathbb{J}^{-1}U_1, U_2 \rangle_{\mathcal{Y}} = \int_{\mathbb{S}^1} (u_1 v_2 - u_2 v_1) dx \quad (3.24)$$

for all $U_1 = (u_1, v_1)^T, U_2 = (u_2, v_2)^T \in \mathcal{Y}$.

Since the Laplacian is diagonal in the Fourier representation (3.17) with eigenvalues $-k^2$ for $k \in \mathbb{Z}$, the eigenvalue problem for A separates into 2×2 eigenvalue problems on each Fourier mode, and $\text{spec } A = i\mathbb{Z} \setminus \{0\}$. Clearly, A is skew-symmetric on \mathcal{Y} with respect to the inner product (3.18). Note that $\mathbb{P}_0 \tilde{A}$ has a Jordan block and is hence included with the nonlinearity B . Thus,

$$\mathcal{Y}_{\tau, \ell} = \mathcal{G}_{\tau, \ell+1}(\mathbb{S}^1; \mathbb{R}) \times \mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{R})$$

with $q = 1$.

The symplectic structure operator \mathbb{J} defined by (3.24) is an unbounded operator on $\mathcal{Y}_{\tau, \ell}$ with domain $\mathcal{Y}_{\tau, \ell+1}$. It is possible, though not necessary for anything which follows, to compute \mathbb{J}^{-1} explicitly. Namely, (3.24) reads

$$\langle \mathbb{J}^{-1}U_1, U_2 \rangle_{\mathcal{Y}} = \langle (\mathbb{J}^{-1}U_1)_u, u_2 \rangle_{\mathcal{H}^1} + \langle (\mathbb{J}^{-1}U_1)_v, v_2 \rangle_{\mathcal{L}^2} = \int_{\mathbb{S}^1} (u_1 v_2 - u_2 v_1) dx. \quad (3.25)$$

The definition of the inner product (3.18) implies

$$\langle (\mathbb{J}^{-1}U_1)_u, u_2 \rangle_{\mathcal{H}^1} = \langle (p_0 - \partial_x^2)(\mathbb{J}^{-1}U_1)_u, u_2 \rangle_{\mathcal{L}^2},$$

so that (3.25) splits into

$$(p_0 - \partial_x^2)(\mathbb{J}^{-1}U_1)_u = -v_1 \quad \text{and} \quad (\mathbb{J}^{-1}U_1)_v = u_1.$$

We conclude that

$$\mathbb{J}^{-1} = \begin{pmatrix} 0 & -(p_0 - \partial_x^2)^{-1} \\ 1 & 0 \end{pmatrix}.$$

If the potential $V: D \rightarrow \mathbb{R}$ is analytic on an open set $D \subset \mathbb{R}$, then, by Lemma 3.10, B is analytic from $\mathcal{B}_R^{\mathcal{Y}_{\tau, \ell}}(U^0)$ to $\mathcal{Y}_{\tau, \ell}$ for any $\tau, \ell \geq 0$ and $U^0 = (u^0, v^0)$ provided $\mathcal{B}_r^{\mathbb{R}}(u^0) \subset D$, $R < r/c(\ell)$ as in Lemma 3.10, and $v^0 \in \mathcal{L}_2$ arbitrary. In this setting, all the above assumptions are satisfied.

3.6. Functional setting for the nonlinear Schrödinger equation. Consider the nonlinear Schrödinger equation

$$i \partial_t u = -\partial_{xx} u + \partial_{\bar{u}} V(u, \bar{u}) \quad (3.26)$$

on the circle \mathbb{S}^1 , where $V(u, \bar{u})$ is analytic in $\text{Re } u$ and $\text{Im } u$. Setting $U \equiv u$, we can write

$$A = i \partial_x^2, \quad B(U) = -i \partial_{\bar{u}} V(u, \bar{u}), \quad (3.27)$$

and, similarly to (3.24),

$$\langle \mathbb{J}^{-1}u_1, u_2 \rangle_{\mathcal{Y}} = \int \text{Re}(iu_1 \bar{u}_2) dx. \quad (3.28)$$

The Hamiltonian is given by

$$H(U) = \frac{1}{2} \int_{\mathbb{S}^1} (|\partial_x u|^2 + V(u, \bar{u})) dx \quad (3.29)$$

for all $u_1, u_2 \in \mathcal{Y} \equiv \mathcal{H}^1(\mathbb{S}^1, \mathbb{C})$. The Laplacian is diagonal in the Fourier representation (3.17) with eigenvalues $-k^2$. Hence, $\text{spec } A = \{-ik^2 : k \in \mathbb{Z}\}$ and A generates a unitary group on $\mathcal{L}^2(\mathbb{S}^1; \mathbb{C})$ and, more generally, on every $\mathcal{G}_{\tau, \ell}$ with $\ell \in \mathbb{N}_0$ and $\tau \geq 0$. Further, $\mathcal{Y}_{\tau, \ell} = \mathcal{G}_{\tau, 2\ell+1}(\mathbb{S}^1; \mathbb{C})$ with $q = 2$.

If the potential $V : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ is analytic as a function of $(q, p) \equiv (\text{Re } u, \text{Im } u)$, then, by Lemma 3.10, the nonlinearity $B(U)$ defined in (3.27) is analytic as map from a ball in $\mathcal{G}_{\tau, \ell}(\mathbb{S}^1; \mathbb{C})$ to itself for every $\tau \geq 0$ and $\ell > 1/2$. The construction of the domain hierarchy works as in Section 3.5, so that all of the above assumptions can be satisfied.

We remark that if we were to write out the nonlinear Schrödinger equation in real coordinates with $U = (\text{Re } u, \text{Im } u)$, the structure operator \mathbb{J} would be the canonical symplectic matrix on \mathbb{R}^2 .

3.7. Nonlocal Schrödinger equation on a star-shaped domain. Consider the nonlocal nonlinear Schrödinger equation

$$i \partial_t u = -\partial_{xx} u + V' \left(\int_0^{2\pi} |u|^2 dx \right) u \quad (3.30)$$

with periodic boundary conditions. It is Hamiltonian with symplectic structure as above and energy

$$H(U) = \frac{1}{2} \int_{\mathbb{S}^1} |\partial_x u|^2 dx + V \left(\frac{1}{2} \int_0^{2\pi} |u|^2 dx \right).$$

In this case,

$$B(u) = V' \left(\int_0^{2\pi} |u|^2 dx \right) u$$

and A is as above. Let, for example, $V(r) = 1/r$. Then B is analytic as map from $\mathcal{Y}_{\tau, \ell} \setminus \{0\}$ to $\mathcal{Y}_{\tau, \ell}$ where $\mathcal{Y}_{\tau, \ell} = \mathcal{G}_{\tau, 2\ell+1}$ as before. Let $\{e_k\}_{k \in \mathbb{Z}}$ denote a \mathcal{Y} -orthonormal basis of eigenvectors of A . Fix $R > 0$ and define

$$\mathcal{D} = \text{int}(\mathcal{B}_R^{\mathcal{Y}}(0)) \setminus \mathcal{C}_- \quad \text{where} \quad \mathcal{C}_- = \{u = \sum_{j \in \mathbb{Z}} u_n e_j : u_0 \leq 0, \|u - u_0 e_0\|_{\mathcal{Y}} \leq |u_0|\}.$$

Then \mathcal{D} is star-shaped with respect to any u in the cone

$$\mathcal{C}_+ \equiv \{u = \sum_{j \in \mathbb{Z}} u_j e_j : u_0 \geq 0, \|u - u_0 e_0\|_{\mathcal{Y}} \leq |u_0|\} \cap \mathcal{B}_R^{\mathcal{Y}}(0).$$

Moreover, for some $\delta_* > 0$ depending on R , $\mathcal{C}_+^{-\delta}$ is nonempty for all $\delta \in (0, \delta_*)$. Fixing a suitable $\delta_* > 0$, $K \in \mathbb{N}$, and defining the sets $\mathcal{D}_0, \dots, \mathcal{D}_K$ as in Example 3.8, we obtain a δ_* -nested domain hierarchy which satisfies (B1) and (H4). If we define $\mathcal{D}^+ \equiv \text{int}(\mathcal{B}_R^{\mathcal{Y}}(0)) \setminus \mathcal{C}_+$ then we can define δ_* -nested domains $\mathcal{D}_0^+, \dots, \mathcal{D}_K^+$ satisfying (B1) and (H4) as above such that $\mathcal{D} \cup \mathcal{D}^+$ is the punctured open disk $\text{int}(\mathcal{B}_R^{\mathcal{Y}}(0)) \setminus \{0\}$.

4. A-STABLE RUNGE-KUTTA METHODS ON HILBERT SPACES

In this section, we introduce a class of A-stable Runge-Kutta methods which are well-defined when applied to the semilinear PDE (1.1) under assumptions (A) and (B0), and review some regularity and convergence results for those methods from [27]. In this section, we need not assume that (1.1) is Hamiltonian.

Applying an s -stage Runge–Kutta method of the form (2.2) to the semilinear evolution equation (1.1), we obtain

$$W = U^0 \mathbb{1} + h \mathbf{a} (AW + B(W)), \quad (4.1a)$$

$$U^1 = U^0 + h \mathbf{b}^T (AW + B(W)). \quad (4.1b)$$

For $U \in \mathcal{Y}$, we write

$$\mathbb{1} U = \begin{pmatrix} U \\ \vdots \\ U \end{pmatrix} \in \mathcal{Y}^s, \quad W = \begin{pmatrix} W^1 \\ \vdots \\ W^s \end{pmatrix}, \quad B(W) = \begin{pmatrix} B(W^1) \\ \vdots \\ B(W^s) \end{pmatrix},$$

where W^1, \dots, W^s are the stages of the Runge–Kutta method,

$$(\mathbf{a}W)^i = \sum_{j=1}^s \mathbf{a}_{ij} W^j, \quad \mathbf{b}^T W = \sum_{j=1}^s \mathbf{b}_j W^j,$$

and A acts diagonally on the stages, i.e., $(AW)^i = AW^i$ for $i = 1, \dots, s$.

A more suitable form, required later, is achieved by rewriting (4.1a) as

$$W = \Pi(W; U, h) \equiv (\text{id} - h\mathbf{a}A)^{-1} (\mathbb{1}U + h\mathbf{a}B(W)) \quad (4.2)$$

and

$$\Psi^h(U) = S(hA)U + h\mathbf{b}^T (\text{id} - h\mathbf{a}A)^{-1} B(W(U, h)), \quad (4.3)$$

where S is the so-called *stability function*

$$S(z) = 1 + z\mathbf{b}^T (\text{id} - z\mathbf{a})^{-1} \mathbb{1}. \quad (4.4)$$

We now make a number of assumptions on the method and its interaction with the linear operator A . First, we assume that the method is A-stable in the sense of [20]. Setting $\mathbb{C}^- = \{z \in \mathbb{C} : \text{Re } z \leq 0\}$, the conditions are as follows.

(RK1) The stability function (4.4) is bounded with $|S(z)| \leq 1$ for all $z \in \mathbb{C}^-$.

(RK2) The $s \times s$ matrices $\text{id} - z\mathbf{a}$ are invertible for all $z \in \mathbb{C}^-$.

We also require two further conditions.

(RK3) The matrix \mathbf{a} is invertible.

(RK4) The method is symplectic, i.e., satisfies (2.10).

Gauss–Legendre Runge–Kutta methods satisfy conditions (RK1–4); see [27, Lemma 3.6] for condition (RK1–3) and [15] for (RK4).

In the following, we also need to refer to a set of key estimates on the linear operators which appear in the formulation (4.2) and (4.3) of the Runge–Kutta method, namely

$$\|(\text{id} - h\mathbf{a}A)^{-1}\|_{\mathcal{Y}^s \rightarrow \mathcal{Y}^s} \leq \Lambda, \quad (4.5a)$$

$$\|h\mathbf{a}A(\text{id} - h\mathbf{a}A)^{-1}\|_{\mathcal{Y}^s \rightarrow \mathcal{Y}^s} \leq 1 + \Lambda, \quad (4.5b)$$

$$\|S(hA)\|_{\mathcal{Y}^s \rightarrow \mathcal{Y}^s} \leq 1 + \sigma h \leq c_S, \quad (4.5c)$$

for all $h \in [0, h_*]$ and constants $\Lambda \geq 1$, $\sigma \geq 0$, and $c_S \geq 1$. These estimates naturally hold true on each rung of our hierarchy of spaces. Their proofs are given in [27, Section 3.2].

A-stable Runge–Kutta methods have the remarkable property that their time- h map is of the same regularity class as the flow of the evolution equation stated in

Theorem 3.2. We state this result as an abbreviated version of [27, Theorem 3.15 and Remark 3.17].

Theorem 4.1 (Regularity of numerical method). *Assume (A), (B1), and (RK1–3). Pick $\delta \in (0, \delta_*]$ such that $\mathcal{D}_{K+1}^{-\delta}$ is nonempty. Then there exists $h_* > 0$ such that the components W^j of the stage vector $W(U, h)$ and numerical method Ψ are of class (3.11) with T_* there replaced by h_* here. Moreover, Ψ and W^j map into \mathcal{D}_K . The bounds on W , Ψ and h_* only depend on the bounds afforded by (B1) and (3.10), on the coefficients of the method, on the constants afforded by (4.5), and on δ .*

Remark 4.2. Even when $B: \mathcal{Y} \rightarrow \mathcal{Y}$ is analytic, the numerical time- h map $\Psi^h(U)$ is not analytic in h unless U and the image of B are restricted to a finite dimensional subspace. Take, for example, the linear Schrödinger equation, i.e. equation (3.26) with $B \equiv 0$, discretized by the implicit mid point rule. Then

$$h \mapsto S(hA) e_k = (\text{id} + \tfrac{1}{2}hA) (\text{id} - \tfrac{1}{2}hA)^{-1} e_k = \frac{1 + \frac{1}{2}hk^2 i}{1 - \frac{1}{2}hk^2 i} e_k$$

has radius of analyticity $|h| \leq \frac{2}{k}$, where e_k is the k -th Galerkin mode of A as described in Section 3.6. Therefore, if the Fourier expansion of U does not terminate finitely, then $\Psi^h(U) = S(hA)U$ cannot be analytic in h . This argument applies to any A-stable Runge–Kutta method: Since $|S(z)| \leq 1$ for all $z \in i\mathbb{R}$ by assumption (RK1), the stability function is a rational polynomial $S(z) = P(z)/Q(z)$ with $\deg Q \geq \deg P$. Hence, $\deg Q \geq 1$ so that $S(z)$ has at least one pole z_0 . The radius of analyticity of $S(z)$ around 0 is $r_0 = |z_0|$, so that $h \mapsto S(hA)e_k$ cannot be analytic outside a ball around $h = 0$ of radius $|h| \leq r_0/k$. As for the implicit midpoint rule, this implies that $\Psi^h(U)$ is not analytic in h unless the Fourier expansion of U is finite.

Thus, while differentiability in h can be obtained by stepping down on a scale of Hilbert spaces, analyticity can only be obtained by projecting onto a subspace on which the vector field is bounded. This will become necessary in Section 5.3 where analyticity is essential for obtaining exponential error estimates.

In [27, Theorem 3.20], we proved convergence of the time semidiscretization. Condition (A2) of [27] is always satisfied in the setting of this paper and stated as (4.5c) above, so that the convergence theorem, which will be needed later on, can be stated as follows.

Theorem 4.3 (Convergence of time semidiscretization). *Assume (A) and apply a Runge–Kutta method of classical order p subject to conditions (RK1–3) to it. Assume further that (B1) holds with $K \geq p$. Pick $\delta \in (0, \delta_*]$ such that $\mathcal{D}_{p+1}^{-\delta}$ is non-empty and fix $T > 0$. Then there exist positive constants h_* , c_1 , and c_2 which only depend on the bounds afforded by (B1) and (3.10), on the coefficients of the method, on the constants from (4.5), and on δ , such that for every U^0 with*

$$\{\Phi^t(U^0) : t \in [0, T]\} \subset \mathcal{D}_{p+1}^{-\delta} \quad (4.6)$$

and for every $h \in [0, h_]$, the numerical solution $(\Psi^h)^n(U^0)$ lies in \mathcal{D} and satisfies*

$$\|(\Psi^h)^n(U^0) - \Phi^{nh}(U^0)\|_{\mathcal{Y}} \leq c_2 e^{c_1 nh} h^p$$

so long as $nh \leq T$.

5. BACKWARD ERROR ANALYSIS FOR TIME-SEMIDISCRETIZATIONS

5.1. Statement of the main result. We are now ready to state and prove our main result on the backward error analysis of semilinear Hamiltonian evolution equations. As before, we write $j = \lfloor r \rfloor$ to denote the largest integer $j \leq r$, and $j = \lceil r \rceil$ to denote the smallest integer $j \geq r$.

Theorem 5.1 (Main theorem). *Assume that the semilinear Hamiltonian evolution equation (3.12) with energy (3.14) satisfies (A), (B0-2), and (H0-4). Apply a symplectic Runge-Kutta method of order $p \geq 1$ and stepsize h which satisfies (RK1-4) to (3.12). Assume further that $K \geq a + b$,*

$$a = \lceil p(q+1) + q \rceil \quad \text{and} \quad b = \lceil p(q+1)/q \rceil \quad (5.1)$$

with q defined in (B2). Then for each sufficiently small $\delta \in (0, \delta_)$ there exists $h_* > 0$ and a modified energy $\tilde{H}: \mathcal{D}_1^{-\delta/2} \times [0, h_*] \rightarrow \mathbb{R}$ which is analytic in U for each $h \in [0, h_*]$ and satisfies*

$$\sup_{U \in \mathcal{D}_{a+b}^{-\delta}} |\tilde{H}(U, h) - H(U, h)| = O(h^p). \quad (5.2)$$

Moreover, the modified energy is approximately conserved with exponentially small error in the sense that there is $c_ > 0$ such that*

$$\sup_{U \in \mathcal{D}_{\tau, L+1}^{-\delta}} |\tilde{H}(\Psi^h(U), h) - \tilde{H}(U, h)| = O(e^{-c_* h^{-\frac{1}{1+q}}}). \quad (5.3)$$

For the semilinear wave equation, $q = 1$ (see Section 3.5), so that the exponent in (5.3) scales like $h^{-1/2}$. In the case of nonlinear Schrödinger equations, $q = 2$ (see Section 3.6), so that the exponent in (5.3) scales like $h^{-1/3}$.

The proof of this theorem will be developed as a series of Lemmas in Sections 5.2–5.4 below, the two claims corresponding to Lemma 5.17 and Lemma 5.10, respectively. Note that, due to (B2), $\mathcal{D}_{\tau, L+1}^{-\delta} \subset \mathcal{D}_{K+1}^{-\delta} \subset \mathcal{D}_{a+b}^{-\delta}$, so that the supremum in (5.2) can be taken, in particular, over $\mathcal{D}_{\tau, L+1}^{-\delta}$.

Remark 5.2. If the entire discrete orbit $\{U^j = (\Psi^h)^j(U^0): j \in \mathbb{N}_0\}$ lies in $\mathcal{D}_{\tau, L+1}^{-\delta}$, then a bound of the form (5.3) holds, in particular, for all $U = U^j$ so that the energy remains approximately conserved over exponentially long times. This condition is, for example, satisfied when the discrete orbit is contained in a finite-dimensional invariant set for Ψ^h which could be obtained by discretizing a periodic orbit or another finite-dimensional invariant set of the semilinear evolution equation (1.1). Kuksin and Pöschel [30] and Pöschel [31], for example, have used KAM methods to prove existence of finite-dimensional invariant tori for the nonlinear Schrödinger equation and the semilinear wave equation, respectively. It is conceivable that such methods could be extended to discrete symplectic maps Ψ^h .

Remark 5.3. It has been shown for certain types of nonlinear Schrödinger equations that solutions with small enough initial data remain analytic over exponentially long times under generic non-resonance conditions [10]. It is conceivable that such results also hold for A-stable Runge-Kutta time semidiscretizations of nonlinear Schrödinger equations. In [12] such results are shown for splitting methods and in [4] for trigonometric integrators applied to the nonlinear Schrödinger equation, but only over polynomially long times.

Remark 5.4. When the evolution equation (1.1) is linear, e.g., a linear wave equation or linear Schrödinger equation, its Hamiltonian is conserved exactly since symplectic Runge–Kutta methods conserve quadratic invariants [15].

When an orbit of the full semilinear evolution equation lies in $\mathcal{D}_{\tau,L+1}^{-\delta}$ over a finite interval of time, consistency of the numerical method implies that the discrete orbit will lie in a slightly larger subset of a Hilbert space of less regularity provided h is sufficiently small. Then the conclusions of Theorem 5.1 hold true at least on this finite interval of time. To make this argument rigorous, we need to invoke our convergence result, Theorem 4.3, with $\mathcal{Y}_{\tau,L+1}$ at the base of the scale, which requires a further assumption.

(B3) There exist $\tau > 0$, $q > 0$, $L \geq 0$, and a sequence of δ_* -nested $\mathcal{Y}_{\tau,L+k}$ -bounded and open sets $\mathcal{D}_{\tau,L+k}$ such that $B \in \mathcal{C}_b^{3-k}(\mathcal{D}_{\tau,L+k}, \mathcal{Y}_{\tau,L+k})$ for $k = 0, 1, 2$.

Further, let $\mathcal{D}_{\tau,L+3}$ denote any δ_* -nested subset of $\mathcal{D}_{\tau,L+2}$ which is bounded and open in $\mathcal{Y}_{\tau,L+3}$.

Corollary 5.5. *Under the assumptions of Theorem 5.1 suppose, in addition, that (B3) holds true. Fix $T > 0$, $\delta > 0$, and $\varepsilon > 0$ such that $\varepsilon + \delta \leq \delta_*$. Then there exists $h_* > 0$ and a modified Hamiltonian $\tilde{H}: \mathcal{D}_{\tau,L+1}^{-\delta} \times [0, h_*] \rightarrow \mathbb{R}$ such that for any U^0 with*

$$\{\Phi^t(U^0): t \in [0, T]\} \subset \mathcal{D}_{\tau,L+3}^{-\delta-\varepsilon} \quad (5.4)$$

and for any $h \in [0, h_*]$,

$$|\tilde{H}(U^0, h) - H(U^0, h)| = O(h^p),$$

the discrete trajectory satisfies $U^j \equiv (\Psi^h)^j(U^0) \in \mathcal{D}_{\tau,L+1}^{-\delta}$, and for fixed $\beta \in (0, c_*)$,

$$|\tilde{H}(U^j, h) - \tilde{H}(U^0, h)| = O(e^{-\beta h^{-\frac{1}{1+q}}})$$

for $j = 0, \dots, \lfloor T/h \rfloor$. The order constants in the two estimates are uniform over all U^0 satisfying (5.4).

Proof. We first apply Theorem 4.3 with $p \equiv 1$, $\mathcal{Y}_{\tau,L+1}$ in place of \mathcal{Y} , and with $\delta + \varepsilon$ in place of δ . We conclude, in particular, that there is $h_* > 0$ such that for any $h \in (0, h_*]$, we have $U^j \in \mathcal{D}_{\tau,L+1}^{-\delta}$ so long as $0 \leq j \leq \lfloor T/h \rfloor$.

Next, Theorem 5.1 asserts that there there is a (possibly decreased) $h_* > 0$ and a modified Hamiltonian $\tilde{H}: \mathcal{D}_{\tau,L+1}^{-\delta} \times [0, h_*] \rightarrow \mathbb{R}$ such that (5.3) holds true, and we estimate

$$|\tilde{H}(U^j) - \tilde{H}(U^0)| \leq \sum_{i=0}^{j-1} |\tilde{H}(U^{i+1}) - \tilde{H}(U^i)| \leq h^{-1} O(e^{-c_* h^{-\frac{1}{1+q}}}) = O(e^{-\beta h^{-\frac{1}{1+q}}})$$

for any $\beta \in (0, c_*)$. □

The remainder of the paper is devoted to the proof of Theorem 5.1. It is based on an embedding result, Lemma 5.8, which generalizes Theorem 2.1 to the Hilbert space setting. However, it is not possible to directly apply the theorems of Section 2, because the formal expansions of both, the numerical method and the modified vectorfield, contain powers of the unbounded operator A . Therefore, the modified vectorfield cannot be written as a semilinear Hamiltonian evolution equation of the form (3.12). If we were simply interested in constructing the modified vectorfield,

we could get around this issue by setting up different spaces for domain and range such that the modified vectorfield, computed up to a given order, is continuous. In fact, we need such techniques to obtain the optimal order estimates for the modified Hamiltonian which is yet to be constructed, see Section 5.4. In such a setting, however, we do not have a theory of local existence of solutions for the modified differential equation so that we cannot obtain a flow.

We thus resort to the following construction. In Section 5.2, we truncate the evolution equation (3.12) to the subspace $\mathbb{P}_m\mathcal{Y}$. Then, in Section 5.3, we obtain a modified flow on this subspace and relax the cut-off m in a controlled way to obtain an embedding result for the time-semidiscretization. Finally, in Section 5.4, we prove estimate (5.2).

5.2. Galerkin truncation. For given $m \in \mathbb{N}$, we define a truncated Hamiltonian evolution equation by restricting the Hamiltonian phase space to the finite dimensional subspace $\mathbb{P}_m\mathcal{Y}$. Since $\nabla H|_{\mathbb{P}_m\mathcal{Y}} = \mathbb{P}_m\nabla H$ and \mathbb{J}^{-1} leaves $\mathbb{P}_m\mathcal{Y}$ invariant by Lemma 3.4, the corresponding restricted evolution equation reads

$$\dot{u}_m = \mathbb{J}\mathbb{P}_m\nabla H(u_m).$$

Thus, setting $f_m = \mathbb{P}_m F$ and $B_m = \mathbb{P}_m B$, we can write

$$\dot{u}_m \equiv f_m(u_m) = Au_m + B_m(u_m). \quad (5.5)$$

We denote the flow of the projected system on $\mathbb{P}_m\mathcal{Y}$ by ϕ_m^t . For convenience, we set $\Phi_m^t = \phi_m^t \circ \mathbb{P}_m$. Similarly, let w_m denote the stage vector, w_m^j for $j = 1, \dots, s$ its components, and ψ_m^h denote the numerical time- h map obtained by applying an s -stage Runge–Kutta method to the projected semilinear evolution equation (5.5), and abbreviate $W_m^j = w_m^j \circ \mathbb{P}_m$ and $\Psi_m^h = \psi_m^h \circ \mathbb{P}_m$.

In [28], we proved that all of the maps above—the truncated flow Φ_m^t and the components of the stage vector W_m^j and the time- h map Ψ_m^h of the truncated system—are of the same class (3.11) as the exact flow Φ^t with m -independent bounds. The precise statement is as follows.

Theorem 5.6 (Regularity of flow and numerical method of projected system). *Assume (A), (B1), and (RK1–3), and choose a $\delta \in (0, \delta_*]$ such that $\mathcal{D}_{K+1}^{-\delta} \neq \emptyset$. Then there are positive T_* , h_* , and m_* such that for every $m \geq m_*$ the flow Φ_m^t is of class (3.11), and the components of the numerical stage vector W_m^j and the numerical time- h map Ψ_m^h are of the same class, but with T_* replaced by h_* , with bounds which are independent of $m \geq m_*$. Moreover, Φ_m^t , W_m^j , and Ψ_m^h map $\mathcal{D}_{K+1}^{-\delta}$ into \mathcal{D}_K .*

Note that Φ_m^t and Ψ_m^h are analytic in t resp. in h so long as B is analytic on \mathcal{Y} . However, the radius of analyticity is generally not uniform in m .

Next, we present an exponential error bound for the projection error for the numerical scheme; this is necessary for obtaining an exponential backward error analysis in Section 5.3 below.

Lemma 5.7 (Exponential projection error estimate for the numerical scheme). *Assume that the semilinear evolution equation (1.1) satisfies conditions (A) and (B0–2). Let, as before, Ψ^h and Ψ_m^h denote a single step of a Runge–Kutta method subject to (RK1–3) applied to the full and the projected semilinear evolution equation, (1.1) and (5.5), respectively. Then for $\delta \in (0, \delta_*]$ there are positive constants*

h_* , m_* , and c_Ψ such that for all $m \geq m_*$, $h \in [0, h_*]$ and $U \in \mathcal{D}_{\tau, L+1}^{-\delta}$,

$$\|\Psi^h(U) - \Psi_m^h(U)\|_{\mathcal{Y}_1} \leq c_\Psi m^{-L} e^{-\tau m^{1/q}}. \quad (5.6)$$

Proof. By Theorems 4.1 and 5.6 applied with $\mathcal{Y}_{\tau, L}$ in place of \mathcal{Y} and $K = 0$, there exist $h_* > 0$ and $m_* > 0$ such that

$$\Psi^h, \Psi_m^h, W^j, W_m^j \in \mathcal{C}_b(\mathcal{D}_{\tau, L+1}^{-\delta} \times [0, h_*]; \mathcal{Y}_{\tau, L+1} \cap \mathcal{D}_{\tau, L})$$

for $j = 1, \dots, s$ with bounds which are uniform in $h \in (0, h_*]$ and $m \geq m_*$.

We first estimate the difference of the stages vectors $W(U) - w_m(U)$, noting that

$$\begin{aligned} W(U) &= (\text{id} - h\mathbf{a}A)^{-1} (\mathbb{1}U + h\mathbf{a}B(W(U))) \\ W_m(U) &= (\text{id} - h\mathbf{a}A)^{-1} (\mathbb{P}_m \mathbb{1}U + h\mathbb{P}_m \mathbf{a}B(w_m(U))). \end{aligned}$$

Taking the difference of both expressions, we obtain

$$W(U) - w_m(U) = (\text{id} - h\mathbf{a}A)^{-1} (\mathbb{Q}_m \mathbb{1}U + h\mathbf{a} [B(W(U)) - \mathbb{P}_m B(W_m(U))]). \quad (5.7)$$

By (4.5a),

$$\begin{aligned} \|h\mathbf{a}(\text{id} - h\mathbf{a}A)^{-1} (B(W(U)) - \mathbb{P}_m B(W_m(U)))\|_{\mathcal{Y}^s} \\ \leq h\Lambda \|\mathbf{a}\| \|B(W(U)) - \mathbb{P}_m B(W_m(U))\|_{\mathcal{Y}^s}. \end{aligned}$$

Using the triangle inequality and the mean value theorem, we further estimate

$$\begin{aligned} \|B(W(U)) - \mathbb{P}_m B(W_m(U))\|_{\mathcal{Y}^s} \\ \leq \|\mathbb{Q}_m B(W(U))\|_{\mathcal{Y}^s} + \|\mathbb{P}_m B(W(U)) - \mathbb{P}_m B(W_m(U))\|_{\mathcal{Y}^s} \\ \leq \|\mathbb{Q}_m B(W(U))\|_{\mathcal{Y}^s} + M'_0 \|W(U) - W_m(U)\|_{\mathcal{Y}^s}. \end{aligned} \quad (5.8)$$

Taking the \mathcal{Y}^s norm of (5.7), using (4.5a), and inserting (5.8), we obtain

$$\begin{aligned} \|W(U) - W_m(U)\|_{\mathcal{Y}^s} &\leq \Lambda \|\mathbb{Q}_m U\|_{\mathcal{Y}^s} + h\|\mathbf{a}\| \Lambda \|\mathbb{Q}_m B(W(U))\|_{\mathcal{Y}^s} \\ &\quad + hM'_0 \|\mathbf{a}\| \Lambda \|W(U) - W_m(U)\|_{\mathcal{Y}^s}. \end{aligned}$$

This proves that, for $h_* < 1/(M'_0 \|\mathbf{a}\| \Lambda)$,

$$\|W(U) - w_m(U)\|_{\mathcal{Y}^s} \leq \Lambda \frac{\|\mathbb{Q}_m U\|_{\mathcal{Y}}}{1 - h_* M'_0 \Lambda \|\mathbf{a}\|} + h_* \|\mathbf{a}\| \Lambda \frac{\|\mathbb{Q}_m B(W(U))\|_{\mathcal{Y}^s}}{1 - h_* M'_0 \Lambda \|\mathbf{a}\|}.$$

We apply (3.5) to the first term on the right and note that, again by (3.5),

$$\|\mathbb{Q}_m B(W(U))\|_{\mathcal{Y}^s} \leq m^{-L} e^{-\tau m^{1/q}} M_{\tau, L}, \quad (5.9)$$

where $M_{\tau, L}$ is the bound for the norm of $B: \mathcal{D}_{\tau, L} \rightarrow \mathcal{Y}_{\tau, L}$ afforded by assumption (B2). This establishes that there exists a constant c_W such that

$$\|W(U) - W_m(U)\|_{\mathcal{Y}^s} \leq c_W m^{-L} e^{-\tau m^{1/q}} \quad (5.10)$$

for all $m \geq m_*$ and $h \in [0, h_*]$. (As in [28], we could obtain a stage vector error bound in the \mathcal{Y}_1^s -norm by applying A onto (5.7) and using (4.5b), but this is not necessary for what follows.)

To estimate the difference between the Runge–Kutta updates, we write them in the form (4.3) such that the respective right hand sides are (uniformly) bounded operators,

$$\begin{aligned} \Psi^h(U) &= S(hA) U + h\mathbf{b}^T (\text{id} - h\mathbf{a}A)^{-1} B(W(U)), \\ \Psi_m^h(U) &= S(hA) \mathbb{P}_m U + h\mathbf{b}^T (\text{id} - h\mathbf{a}A)^{-1} \mathbb{P}_m B(w_m(U)). \end{aligned}$$

Then,

$$\Psi^h(U) - \Psi_m^h(U) = S(hA) \mathbb{Q}_m U + h \mathbf{b}^T (\text{id} - h\mathbf{a}A)^{-1} [B(W(U)) - \mathbb{P}_m B(w_m(U))].$$

Inserting (5.9) and (5.10) back into (5.8), we also find that

$$\|B(W(U)) - \mathbb{P}_m B(w_m(U))\|_{\mathcal{Y}^s} \leq c_B m^{-L} e^{-\tau m^{1/q}} \quad (5.11)$$

for some constant $c_B > 0$. We note that (3.4) and (4.5b) imply

$$\|h\mathbf{a}(\text{id} - h\mathbf{a}A)^{-1}\|_{\mathcal{Y}^s \rightarrow \mathcal{Y}_1^s} \leq 2 + \Lambda.$$

This and the invertibility of \mathbf{a} , assumption (RK3), then yield

$$\begin{aligned} & \|\Psi^h(U) - \Psi_m^h(U)\|_{\mathcal{Y}_1} \\ & \leq (1 + \sigma h) \|\mathbb{Q}_m U\|_{\mathcal{Y}_1} + \|\mathbf{b}\| \|\mathbf{a}^{-1}\| \|h\mathbf{a}(\text{id} - h\mathbf{a}A)^{-1}(B(W(U)) - \mathbb{P}_m B(w_m(U)))\|_{\mathcal{Y}_1^s} \\ & \leq (1 + \sigma h) \|\mathbb{Q}_m U\|_{\mathcal{Y}_1} + (2 + \Lambda) \|\mathbf{b}\| \|\mathbf{a}^{-1}\| \|B(W(U)) - \mathbb{P}_m B(w_m(U))\|_{\mathcal{Y}^s}. \end{aligned}$$

Inequality (5.6) is now a consequence of (3.5) and (5.11). \square

5.3. Embedding of semidiscretizations into a flow. We now apply the backward error analysis of Section 2 to the sequence of truncated problems defined above where, for each m , we work on the finite dimensional space $\mathbb{P}_m \mathcal{Y}$ endowed with the \mathcal{Y}_1 -norm.

By construction, f_m is analytic on $\mathcal{D} \cap \mathbb{P}_m \mathcal{Y}$ for each $m \in \mathbb{N}$ and satisfies the estimate

$$\begin{aligned} \|f_m(u_m)\|_{\mathcal{Y}_1} & \leq \|A u_m\|_{\mathcal{Y}_1} + \|B_m(u_m)\|_{\mathcal{Y}_1} \\ & \leq \sup_{U \in \mathcal{D}_1} \|A \mathbb{P}_m U\|_{\mathcal{Y}_1} + m \sup_{U \in \mathcal{D}} \|\mathbb{P}_m B(U)\|_{\mathcal{Y}} \leq c_F m \end{aligned} \quad (5.12)$$

for $u_m \in \mathcal{D} \cap \mathbb{P}_m \mathcal{D}_1$ with

$$c_F = \sup_{U \in \mathcal{D}_1} \|U\|_{\mathcal{Y}_1} + M_0,$$

where the supremum is finite as \mathcal{D}_1 is bounded.

Setting $M = c_F m$, Theorem 2.1 asserts that the numerical time- h map can be embedded into a modified flow up an error which is exponentially small in the step size h , albeit not uniformly in m . If, however, we make the stronger assumption that the initial data lies some Gevrey space $\mathcal{Y}_{\tau, L+1}$ with $\tau > 0$, Lemma 5.7 asserts that the numerical solution to the full semilinear evolution equation (1.1) remains exponentially close in the spectral cutoff m to the numerical solution of the projected system. Thus, we can carefully choose $m = m(h)$ to balance the projection error and the embedding error bounds to obtain an embedding result on the Gevrey space which is still exponential in h , but at a lesser rate. This is done in the next lemma, where we also show that the result can be formulated not only on balls as in Theorem 2.1, but also on a more general m -independent subdomains of \mathcal{Y}_1 as needed in the proof of Theorem 5.1.

Following the notation of Section 2, we denote the coefficients of the power series expansion of ψ_m^h by g_m^j , the expansion coefficients of the modified vector field—defined via (2.5)—by f_m^j , and seek an optimally truncated modified vector field of the form

$$\tilde{f}_m^n(u_m; h) = f_m(u_m) + \sum_{j=p}^{n-1} h^j f_m^{j+1}(u_m). \quad (5.13)$$

Lemma 5.8 (Embedding lemma for Gevrey class data). *Assume that the semilinear evolution equation (1.1) satisfies conditions (A) and (B0–2). Let, as before, Ψ^h and ψ_m^h denote a single step of a Runge–Kutta method subject to (RK1–3) applied to the full and the projected semilinear evolution equation, (1.1) and (5.5), respectively. Then for $\delta \in (0, \delta_*)$ there exists $h_* > 0$ such that the choices*

$$m(h) = \left(\frac{\chi}{\tau h} \right)^{\frac{q}{1+q}} \quad \text{and} \quad n(h) = \left\lfloor \tau^{\frac{q}{1+q}} \left(\frac{\chi}{h} \right)^{\frac{1}{1+q}} / 4 \right\rfloor \quad (5.14)$$

with $\chi = \delta / (2\eta c_F)$ ensure that the modified vector field

$$\tilde{F}(U; h) \equiv \tilde{f}_{m(h)}^{n(h)}(\mathbb{P}_m U; h), \quad (5.15)$$

where \tilde{f}_m^n is given by (5.13), is well defined as an analytic map from $\mathcal{D}_1^{-\delta/2}$ to \mathcal{Y}_1 for every fixed $h \in [0, h_*]$ and generates a modified flow $\tilde{\Phi}: \mathcal{D}_1^{-\delta} \times [0, h_*] \rightarrow \mathcal{D}_1^{-\delta/2}$.

Moreover, for every $c_* \in (0, \tau^{\frac{q}{1+q}} \chi^{\frac{1}{1+q}})$ there exists a constant $c_{\tilde{\Phi}}$ such that for every $U \in \mathcal{D}_{\tau, L+1}^{-\delta}$ and $h \in [0, h_*]$

$$\|\Psi^h(U) - \tilde{\Phi}^h(U)\|_{\mathcal{Y}_1} \leq c_{\tilde{\Phi}} e^{-c_* h^{-\frac{1}{1+q}}}. \quad (5.16)$$

Proof. Set $r = \delta/4$. As \mathcal{D}_1 is a bounded subset of \mathcal{Y}_1 , there exists m_* such that, due to (3.5), $\|\mathbb{Q}_m U\|_{\mathcal{Y}} \leq m^{-1} \|U\|_{\mathcal{Y}_1} \leq \delta/2 - r$ and therefore $\mathbb{P}_m U \in \mathcal{D}^{-r}$ for every $U \in \mathcal{D}_1^{-\delta/2}$ and $m \geq m_*$. In particular, for any such U and m ,

$$\mathcal{B}_r^{\mathbb{P}_m \mathcal{Y}_1}(\mathbb{P}_m U) = \{u \in \mathbb{P}_m \mathcal{Y}: \|u - \mathbb{P}_m U\|_{\mathcal{Y}_1} \leq r\}$$

is contained in $\mathbb{P}_m \mathcal{D}_1 \cap \mathcal{D}$ such that estimate (5.12) holds true and we can apply Theorem 2.1 with $M = c_F m$ on this ball. This theorem asserts that for every $h \in [0, h_0/4]$, where $h_0 = r/(2\eta c_F m)$, the modified vector field $\tilde{f}_m = \tilde{f}_m^{n(h)}$ with $n(h) = \lfloor h_0/h \rfloor$ is defined on $\mathcal{B}_{r/4}^{\mathbb{P}_m \mathcal{Y}_1}(\mathbb{P}_m U)$ and analytic as map from $\mathcal{B}_{r/4}^{\mathbb{P}_m \mathcal{Y}_1}(\mathbb{P}_m U)$ to \mathcal{Y}_1 . Its flow $\tilde{\phi}_m^t$ satisfies $\tilde{\phi}_m^t(\mathbb{P}_m U) \in \mathcal{B}_{r/4}^{\mathbb{P}_m \mathcal{Y}_1}(\mathbb{P}_m U)$ for at least $0 \leq t \leq h$, and

$$\|\psi_m^h(\mathbb{P}_m U) - \tilde{\phi}_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} \leq h \gamma c_F m e^{-h_0/h}. \quad (5.17)$$

By construction, $\tilde{F}_m = \tilde{f}_m \circ \mathbb{P}_m$ is analytic as map from $\mathcal{D}_1^{-\delta/2}$ to \mathcal{Y}_1 and $\tilde{\Phi}_m = \tilde{\phi}_m \circ \mathbb{P}_m$ is analytic as map from $\mathcal{D}_1^{-\delta}$ to $\mathcal{D}_1^{-\delta/2}$ for any choice of $m \geq m_*$.

Our next step is to estimate the difference between the solution to the modified projected equation and the numerical solution of the full semilinear evolution equation (1.1). We split the error into the projection and the embedding error, the first of which is controlled by Lemma 5.7 and the second is controlled by (5.17). By Theorems 4.1 and 5.6, respectively, there are $h_* > 0$ and (a possibly increased choice of) m_* such that, for $h \in [0, h_*]$ and $m \geq m_*$, Ψ^h and Ψ_m^h are continuous maps from $\mathcal{D}_{\tau, L+1}^{-\delta}$ to $\mathcal{Y}_{\tau, L+1}$ and such that the exponential projection error estimate asserted by Lemma 5.7 holds. We might have to increase m_* further to ensure that $h_0(m)/4 \leq h_*$, so that the embedding error estimate (5.17) holds true, each for every $m \geq m_*$ and $h \in [0, h_0(m)]$.

Since $\mathcal{D}_{\tau, L+1}^{-\delta} \subset \mathcal{D}_1^{-\delta}$, splitting the total error into a projection error component and the embedding error on the finite dimensional subspace $\mathbb{P}_m \mathcal{Y}$, we obtain that

$$\begin{aligned} \|\Psi^h(U) - \tilde{\phi}_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} &\leq \|\Psi^h(U) - \psi_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} + \|\psi_m^h(\mathbb{P}_m U) - \tilde{\phi}_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} \\ &\leq c_{\Psi} m^{-\ell} e^{-\tau m^{1/q}} + h \gamma c_F m e^{-h_0/h} \end{aligned}$$

for all $U \in \mathcal{D}_{\tau, L+1}^{-\delta}$, $h \in [0, h_*]$, and $m \geq m_*$. The first error decreases with m whereas the second error increases with m . We now demand that the two exponents on the right coincide. Under the ansatz $m = \zeta h^{-\alpha}$ for some ζ and $\alpha \in (0, 1)$, we obtain

$$\|\Psi^h(U) - \tilde{\phi}_m^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} \leq c_\Psi h^{\alpha\ell} \zeta^{-\ell} e^{-\tau \zeta^{1/q} h^{-\alpha/q}} + \gamma c_F \zeta h^{1-\alpha} e^{-\chi \zeta^{-1} h^{\alpha-1}} \quad (5.18)$$

with $\chi = \delta/(2e\eta c_F)$. Then the exponents coincide provided $\tau \zeta^{1/q} h^{-\alpha/q} = \chi \zeta^{-1} h^{\alpha-1}$, i.e., when

$$\alpha = \frac{q}{1+q} \quad \text{and} \quad \zeta = \left(\frac{\chi}{\tau}\right)^\alpha. \quad (5.19)$$

This implies that $m(h)$ is given by (5.14), and

$$\|\Psi^h(U) - \tilde{\phi}_{m(h)}^h(\mathbb{P}_m U)\|_{\mathcal{Y}_1} \leq c_{\tilde{\Phi}} h^\nu e^{-ch^{-\frac{1}{1+q}}} \quad (5.20)$$

with $c = \tau \zeta^{1/q} = \tau^{\frac{q}{1+q}} \chi^{\frac{1}{q+1}}$ and $\nu = \min\{1, q\ell\}/(1+q)$, and where we possibly need to shrink $h_* > 0$ to satisfy

$$m(h_*) \geq m_* \quad \text{and} \quad h_* \leq \frac{\chi \tau^q}{4q+1}.$$

Since $h_0 = \chi m^{-1} = \chi^{1-\alpha} \tau^\alpha h^\alpha$, the latter inequality then guarantees that $h \leq h_0/4$ and yields the form of $n(h)$ stated in (5.14). The exponential estimate (5.16) is then obtained by defining the modified vectorfield by (5.15) with corresponding modified flow $\tilde{\Phi}^t(U) \equiv \tilde{\phi}_{m(h)}^t(\mathbb{P}_m U)$ and dominating the algebraic prefactor by fractional exponential decay. \square

Remark 5.9. The reason for requiring data in $\mathcal{D}_{\tau, L+1}^{-\delta}$ rather than $\mathcal{D}_{\tau, L}^{-\delta}$ is that we want to define Ψ^h and ψ_m^h uniformly in h and m on general open sets, not just on open balls. This can only be done when constructing them as maps $\mathcal{D}_{\tau, L+1}^{-\delta} \rightarrow \mathcal{Y}_{\tau, L+1} \cap \mathcal{D}_{\tau, L}$, see [27, 28]. Moreover, for general domains we can only maintain a valid domain of definition of the nonlinearity under Galerkin truncation and, in particular, assert estimate (5.12) only by dropping down at least one rung on the scale of spaces.

Next, we show that in the Hamiltonian case the above construction also yields a modified Hamiltonian which is approximately conserved under the numerical time- h map of the full semilinear evolution equation (3.12).

Lemma 5.10 (Modified Hamiltonian for Gevrey class data). *Under the conditions and in the notation of Lemma 5.8, suppose further that (H0–4) and (RK4) hold true. Then there exists a modified Hamiltonian $\tilde{H}: \mathcal{D}_1^{-\delta/2} \times [0, h_*] \rightarrow \mathbb{R}$, defined up to a constant of integration, which is analytic in $U \in \mathcal{D}_1^{-\delta/2}$ for every $h \in [0, h_*]$ and such that the modified vectorfield from Lemma 5.8 satisfies $\tilde{F} = \mathbb{J} \nabla \tilde{H}$. Moreover, there exist constants $c_*, c_{\tilde{H}} > 0$ such that for every $U \in \mathcal{D}_{\tau, L+1}^{-\delta}$ and $h \in [0, h_*]$,*

$$|\tilde{H}(\Psi^h(U), h) - \tilde{H}(U, h)| \leq c_{\tilde{H}} e^{-c_* h^{-\frac{1}{1+q}}}. \quad (5.21)$$

Proof. Let $h \in [0, h_*]$ fixed. On each Galerkin subspace $\mathbb{P}_m \mathcal{Y}$, the situation is as in Section 2.3, i.e., the operator $\mathbb{J}^{-1} \text{D}f_m^j(u_m)$ is self-adjoint with respect to the restriction of the \mathcal{Y}_1 -inner product to $\mathbb{P}_m \mathcal{Y}$ for each $u_m \in \mathcal{D} \cap \mathbb{P}_m \mathcal{Y}$. As we argued in the proof of Lemma 5.8, $\mathbb{P}_m U \in \mathcal{D}$ for $U \in \mathcal{D}_1^{-\delta/2}$, so that $\mathbb{J}^{-1} \text{D}\tilde{F}(U)$ is self-adjoint with respect to the \mathcal{Y}_1 -inner product for each $U \in \mathcal{D}_1^{-\delta/2}$. By assumption (H4)

and Theorem 3.6, the set $\mathcal{D}_1^{-\delta/2}$ is simply connected and star-shaped. Therefore Lemma 3.5 on the integrability of vector fields on Hilbert spaces applies on the domain $\mathcal{D}_1^{-\delta/2}$ endowed with the \mathcal{Y}_1 -topology. This proves the existence of an analytic modified Hamiltonian $\tilde{H}: \mathcal{D}_1^{-\delta/2} \rightarrow \mathbb{R}$ which is invariant under the modified flow such that

$$\tilde{F} = \mathbb{J} \nabla \tilde{H}.$$

To prove (5.21), we decrease h_* such that the right hand side of (5.16) is smaller than $r/4$ for every $U \in \mathcal{D}_{\tau, L+1}^{-\delta}$. Then, using the mean value theorem, the bound on the modified vector field given by (2.6) with $M = c_F m$, (5.20) from the proof of Lemma 5.8, and the invertibility of \mathbb{J}^{-1} , we estimate

$$\begin{aligned} |\tilde{H}(\Psi^h(U)) - \tilde{H}(\tilde{\Phi}^h(U))| &\leq \sup_{\substack{y \in \mathbb{P}_m \mathcal{Y} \\ \|y - \mathbb{P}_m U\|_{\mathcal{Y}_1} \leq \frac{r}{4}}} \|\mathrm{D}\tilde{H}(y)\|_{\mathcal{Y}_1} \|\Psi^h(U) - \tilde{\Phi}^h(U)\|_{\mathcal{Y}_1} \\ &\leq \sup_{\substack{y \in \mathbb{P}_m \mathcal{Y} \\ \|y - \mathbb{P}_m U\|_{\mathcal{Y}_1} \leq \frac{r}{4}}} \|\mathbb{J}^{-1} \tilde{f}_m(y)\|_{\mathcal{Y}_1} \|\Psi^h(U) - \tilde{\Phi}^h(U)\|_{\mathcal{Y}_1} \\ &\leq \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y}_1)} (1 + 1.65\eta) c_F m h^\nu c_{\tilde{\Phi}} e^{-ch^{-(1+q)}}. \end{aligned}$$

Now we note that \mathbb{J} and A commute, see (3.13), which, as $\mathbb{J}^{-1} \in \mathcal{E}(\mathcal{Y})$ by (H0), implies $\mathbb{J}^{-1} \mathcal{E}(\mathcal{Y}_1)$. Since \tilde{H} is conserved under the modified flow, choosing m as in (5.14), we obtain

$$|\tilde{H}(\Psi^h(U)) - \tilde{H}(U)| \leq \tilde{c} h^{\nu - \frac{q}{1+q}} e^{-ch^{-\frac{1}{1+q}}}.$$

This inequality implies (5.21) with $c_* > 0$ possibly decreased from its value in Lemma 5.8. \square

What is still missing is a statement regarding the closeness of the modified to the original Hamiltonian. Ideally, we would like H and \tilde{H} to be $O(h^p)$ -close. In a first attempt to prove such a result, we write

$$|H(U) - \tilde{H}(U)| \leq |H(U) - H(\mathbb{P}_m U)| + |H(\mathbb{P}_m U) - \tilde{H}(U)|, \quad (5.22)$$

where $m = m(h)$ is as in (5.14). Under the assumptions of Lemma 5.10, the first term on the right is exponentially small for $U \in \mathcal{D}_{\tau, L+1}$. Indeed, noticing that m_* and hence also $m(h) \geq m_*$ in the proof of Lemma 5.8 were chosen such that, in particular, $\mathbb{P}_m U \in \mathcal{D}$ holds for all $U \in \mathcal{D}_{\tau, L+1}^{-\delta}$, we can estimate

$$\begin{aligned} |H(U) - H(\mathbb{P}_m U)| &\leq \frac{1}{2} |\langle \mathbb{Q}_m U, \mathbb{J}^{-1} A \mathbb{Q}_m U \rangle| + |V(U) - V(\mathbb{P}_m U)| \\ &\leq \|\mathbb{J}^{-1} A\|_{\mathcal{E}(\mathcal{Y})} \|\mathbb{Q}_m U\|_{\mathcal{Y}}^2 + \max_{U \in \mathcal{D}} \|\nabla V(U)\|_{\mathcal{Y}} \|\mathbb{Q}_m U\|_{\mathcal{Y}} \\ &\leq \|\mathbb{J}^{-1} A\|_{\mathcal{E}(\mathcal{Y})} (\|\mathbb{Q}_m U\|_{\mathcal{Y}} + \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y})} M_0) \|\mathbb{Q}_m U\|_{\mathcal{Y}} \\ &\leq c_H \|\mathbb{Q}_m U\|_{\mathcal{Y}} \leq c_H m^{-L-1} e^{-\tau m^{1/q}} \|U\|_{\mathcal{Y}_{\tau, L+1}}. \end{aligned} \quad (5.23)$$

In the first inequality above, we used that \mathbb{J}^{-1} and \mathbb{P}_m commute, cf. Lemma 3.4. The second inequality is based on the mean value theorem and the third inequality holds due to the identity $\nabla V(U) = \mathbb{J}^{-1} B(U)$. The constant c_H is finite due to the boundedness of the operators $\mathbb{J}^{-1} A$ and \mathbb{J}^{-1} provided by (H0), and the boundedness of the domain $\mathcal{D}_{\tau, \ell+1}$. With (5.14), the final line in (5.23) also exponentially small with respect to h .

To estimate the second term of (5.22) choose some fixed $U^0 \in \text{ck}(\mathcal{D}_1)^{-\delta}$ and set $H(U^0) = \tilde{H}(U^0)$. The naive choice is then to employ Lemma 2.2, with $u_m \equiv \mathbb{P}_m U$, $u_m^0 \equiv \mathbb{P}_m U^0$, $M = c_F m$ where m is given by (5.14), and $r = \delta/4$ as in the proof of Lemma 5.8. This yields

$$\begin{aligned} |H(u_m) - \tilde{H}_m(u_m)| &\leq \left| \int_0^1 \langle u_m - u_m^0, \mathbb{J}^{-1}(f_m - \tilde{f}_m)(u_m^0 + s(u - u_m^0)) \rangle ds \right| \\ &\leq \frac{c(\eta, p)}{4} \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y}_1)} r^{-p} M^p \sup_{U \in \mathcal{D}_1} \|U - U^0\|_{\mathcal{Y}_1} h^p \\ &= O(h^{\frac{p-q}{q+1}}) \end{aligned} \quad (5.24)$$

for every $U \in \mathcal{D}_1^{-\delta}$. In the final step, we used that \mathbb{J} and A commute so that $\|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y}_1)} = \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y})}$. Gevrey class regularity does not improve this part of the estimate.

Clearly, (5.24) is much worse than the expected $O(h^p)$. A superficial improvement might be obtained by “paying” with the exponent in (5.21): for any $\varepsilon > 0$, setting $\alpha = \varepsilon/(p+1)$ in (5.19) will yield $O(h^{p-\varepsilon})$ in (5.24).

A closer inspection reveals that, in the context of the semilinear evolution equation, the second inequality of (2.8) in the proof of Lemma 2.2 is too weak: when estimating the terms f^j of the modified vectorfield in some fixed Hilbert space norm, the unboundedness of the operators A contained therein will introduce a factor m^j , see (2.7). This propagates into the proof of (5.24). Note, however, that these estimates are simply about consistency, not about constructing a flow. Thus, we can afford to lose smoothness rather than order. In other words, we can estimate the f^j as maps from one Hilbert space into another with a weaker norm. This will be detailed in the next section.

5.4. Backward error analysis on Hilbert spaces. In this section, we present a more subtle estimate on the difference between the original Hamiltonian and the modified Hamiltonian. The main difference to the derivation of (5.24) in the previous section is that we consider the expansion coefficients of the numerical flows and of the modified vector fields as maps between different rungs on our scale of Hilbert spaces such that the “loss of derivatives” is carefully accounted for.

We begin by establishing the necessary functional setting for the backward error analysis on Hilbert spaces. We then review a result from [28] on the Galerkin projection error for the numerical time- h maps. This estimate is then propagated into an estimate on the difference between the full and the modified vector field, which finally implies a corresponding estimate on the difference between the exact Hamiltonian and the modified Hamiltonian.

In this section, we work directly with the standard construction of the modified vector field. Namely, for $\ell = 1, \dots, K+1$, we write

$$G^\ell = \left. \frac{\partial_h^\ell \Psi^h}{\ell!} \right|_{h=0} \quad (5.25)$$

to denote the ℓ th coefficient of the expansion of Φ^h in powers of h , and define the corresponding expansion coefficients for the modified vector field on Hilbert spaces,

setting $F^1 \equiv G^1$ and

$$F^\ell = G^\ell - \sum_{i=2}^{\ell} \frac{1}{i!} \sum_{\ell_1 + \dots + \ell_i = \ell} D_{\ell_1} \cdots D_{\ell_{i-1}} F^{\ell_i}, \quad (5.26)$$

for $\ell = 2, \dots, K+1$, where the sum ranges over indices $\ell_i \geq 1$ for all i , and where $D_j G = DG F^j$. We also recall from Section 5.3 that g_m^ℓ and f_m^ℓ denote the ℓ th coefficients of the expansions of the projected flow and the corresponding modified vector field, respectively, and set $G_m^\ell \equiv g_m^\ell \circ \mathbb{P}_m$ and $F_m^\ell \equiv f_m^\ell \circ \mathbb{P}_m$.

In the notation of condition (B1), we fix $\delta \in (0, \delta_*]$ and abbreviate $\mathcal{U}_\kappa = \mathcal{D}_\kappa^{-\delta}$ for $\kappa = 1, \dots, K+1$. Then the regularity results Theorem 4.1 on Ψ^h and Theorem 5.6 on Ψ_m^h imply that, in particular, there exists m_* such that for all $m \geq m_*$ and $\ell = 1, \dots, K+1$,

$$G^\ell, G_m^\ell \in \bigcap_{\substack{j+k \leq N \\ \ell \leq k \leq K+1}} \mathcal{C}_b^j(\mathcal{U}_\kappa; \mathcal{Y}_{k-\ell}). \quad (5.27)$$

Moreover, bounds in the norms associated with (5.27) are uniform in $m \geq m_*$.

In the following, we will state such bounds on vector fields in terms of the three-parameter family of norms

$$|g|_{N,K,S} = \max_{\substack{j+k \leq N \\ S \leq k \leq K}} \|D^j g\|_{\mathcal{C}(\mathcal{U}_\kappa; \mathcal{E}^j(\mathcal{Y}_k, \mathcal{Y}_{k-S}))} \quad (5.28)$$

for $1 \leq S \leq K \leq N$. The parameter S plays the role of a “loss of derivatives” index as it forces the image of the map be estimated at least S rungs down the scale. We can then prove a simple result on the regularity of the expansion coefficients of the modified vector field.

Lemma 5.11 (Modified vector field on a scale of Banach spaces). *Assume that G^1, \dots, G^{K+1} are of class (5.27). Then the vector fields F^1, \dots, F^{K+1} defined by (5.26) are also of class (5.27).*

Proof. The proof is based on the simple fact that $F \in \mathcal{C}_b^n(\mathcal{Y}_{i+j}, \mathcal{Y}_j)$ and $G \in \mathcal{C}_b^{n+1}(\mathcal{Y}_j, \mathcal{Y})$ implies that $DG F \in \mathcal{C}_b^n(\mathcal{Y}_{i+j}, \mathcal{Y})$. Thus, it remains to observe that repeated application to the terms in the inner sum of (5.26) causes all loss indices to always sum up to ℓ . We proceed by induction in ℓ . The case $\ell = 1$ does not require proof. Assume therefore that $\ell > 1$ and the lemma is proved up to index $\ell - 1$. For $\ell > \ell_1 \geq 1$, we estimate, with $G \equiv D_{\ell_2} \dots D_{\ell_{i-1}} F^{\ell_i}$ that

$$\begin{aligned} |D_{\ell_1} G|_{N,K+1,\ell} &= |DG F^{\ell_1}|_{N,K+1,\ell} \\ &= \max_{\substack{j+k \leq N \\ \ell \leq k \leq K+1}} \|D^j (DG F^{\ell_1})\|_{\mathcal{C}(\mathcal{U}_\kappa; \mathcal{E}^j(\mathcal{Y}_k, \mathcal{Y}_{k-\ell}))} \\ &\leq 2^N \max_{\substack{j+k \leq N \\ \ell \leq k \leq K+1}} \|D^{j+1} G\|_{\mathcal{C}(\mathcal{U}_{\kappa-\ell_1}; \mathcal{E}^{j+1}(\mathcal{Y}_{k-\ell_1}, \mathcal{Y}_{k-\ell}))} \max_{\substack{j+k \leq N \\ \ell_1 \leq k \leq K+1}} \|D^j F^{\ell_1}\|_{\mathcal{C}(\mathcal{U}_\kappa; \mathcal{E}^j(\mathcal{Y}_k, \mathcal{Y}_{k-\ell_1}))} \\ &= 2^N \max_{\substack{j+k \leq N+1-\ell_1 \\ \ell-\ell_1 \leq k \leq K+1-\ell_1}} \|D^j G\|_{\mathcal{C}(\mathcal{U}_\kappa; \mathcal{E}^j(\mathcal{Y}_k, \mathcal{Y}_{k-(\ell-\ell_1)}))} |F^{\ell_1}|_{N,K+1,\ell_1} \\ &\leq 2^N |G|_{N,K+1-\ell_1,\ell-\ell_1} |F^{\ell_1}|_{N,K+1,\ell_1}, \end{aligned} \quad (5.29)$$

provided that G is of class (5.27) with K replaced by $K - \ell_1$ and ℓ replaced by $\ell - \ell_1$. Here the first inequality is based on the product rule and selective weakening of the norm on the domain spaces, thereby increasing the respective operator norms.

The identity between the third and the fourth line is achieved by redefining $j + 1$ as j and $k - \ell_1$ as k . The final inequality holds because $\ell_1 \geq 1$, so that we are strictly extending the range of the running indices. We note that the second term in the final line of (5.29) is bounded by the induction hypothesis. The estimation of the first term in the final line of (5.29) can now be done recursively to resolve the entire product $D_{\ell_1} \cdots D_{\ell_{i-1}} F^{\ell_i}$ from the inner sum of (5.26) in terms of quantities which are bounded by the induction hypothesis. This is always possible, because at the k th step of this process we lose ℓ_k derivatives, and the sum of the loss indices satisfies $\ell_1 + \cdots + \ell_i = \ell$ by construction. \square

We now aim to derive estimates on the difference between F^ℓ and F_m^ℓ with respect to the same type of norm. In [28], we already obtained a related result on the difference between Ψ^h and Ψ_m^h which we can start from. Setting $\mathcal{I} = (0, h_*)$ and $\mathcal{U} = \mathcal{D}_{K+1}^{-\delta}$, we define the norm

$$\|\Psi\|_{N,K} = \max_{\substack{j+k \leq N \\ \ell \leq k \leq K}} \|D_U^j \partial_h^\ell \Psi\|_{\mathcal{L}^\infty(\mathcal{U} \times \mathcal{I}; \mathcal{E}^j(\mathcal{Y}_{K+1}; \mathcal{Y}_{k-\ell}))}$$

for $0 \leq K \leq N$. Then the stability of the numerical method on a scale of Hilbert spaces under spectral truncation can be formulated as follows.

Lemma 5.12 ([28, Theorem 3.7]). *Assume (A), (B1), and (RK1–3). Choose $\delta \in (0, \delta_*]$ small enough such that $\mathcal{D}_{K+1}^{-\delta}$ is nonempty. Then there is $h_* > 0$ such that for every $0 \leq S \leq K$,*

$$\|\Psi^h - \Psi_m^h\|_{N-1-S, K-S} = O(m^{-S}) \quad (5.30)$$

as $m \rightarrow \infty$. The order constants depend only on the bounds afforded by (B1), (3.10), on the coefficients of the method, and on δ .

Remark 5.13. If $N > K + 1$, the statement as proved in [28] actually holds true with $K + 1$ in place of K . It is easier, however, to proceed with the slightly weaker version stated here.

We note that Lemma 5.12 above already provided us with an exponential estimate on $\Psi^h - \Psi_m^h$ for Gevrey regular data, whereas Lemma 5.12 here also asserts bounds on derivatives with respect to h ; the proof is correspondingly more complicated even in spaces of finite order and can be found in [28].

To proceed, we observe that Lemma 5.12 holds with K in the statement of the lemma replaced by any κ between S and K with K as defined in condition (B1). Then, specializing to the particular value $k = \kappa - S$ in the definition of the norm appearing in (5.30), we obtain

$$\max_{\substack{j+\kappa \leq N-1 \\ S+\ell \leq \kappa}} \|D_U^j \partial_h^\ell (\Psi^h - \Psi_m^h)\|_{\mathcal{L}^\infty(\mathcal{U}_\kappa \times \mathcal{I}; \mathcal{E}^j(\mathcal{Y}_\kappa; \mathcal{Y}_{\kappa-S-\ell}))} = O(m^{-S}).$$

Thus, taking the maximum over the allowed range $\kappa \in S, \dots, K$,

$$\max_{\substack{j+\kappa \leq N-1 \\ S+\ell \leq \kappa \leq K}} \|D_U^j \partial_h^\ell (\Psi^h - \Psi_m^h)\|_{\mathcal{L}^\infty(\mathcal{U}_\kappa \times \mathcal{I}; \mathcal{E}^j(\mathcal{Y}_\kappa; \mathcal{Y}_{\kappa-S-\ell}))} = O(m^{-S}). \quad (5.31)$$

Due to the definition of G^ℓ in (5.25), this directly implies that

$$|G^\ell - G_m^\ell|_{N-1, K, S+\ell} = O(m^{-S}) \quad (5.32)$$

for every $\ell \in 1, \dots, K - S$.

Lemma 5.14 (Stability of the modified vector field). *Suppose G^1, \dots, G^K and $\bar{G}^1, \dots, \bar{G}^K$ are of class (5.27), and let F^ℓ and \bar{F}^ℓ denote expansion coefficients of the respective associated modified vector fields defined via (5.26). Then, for $S \in 0, \dots, K$ and every $\ell \in 1, \dots, K - S$,*

$$|F^\ell - \bar{F}^\ell|_{N-1, K, S+\ell} \leq c \max_{1 \leq k \leq \ell} |G^k - \bar{G}^k|_{N-1, K, S+k},$$

where the constant c depends on the bounds on G^k and \bar{G}^k in the norm $|\cdot|_{N-1, K, k}$ for $k = 1, \dots, \ell$.

Proof. The proof follows the same route as the proof of Lemma 5.11. We set $\bar{D}_\ell G \equiv DG\bar{F}_\ell$. The crucial estimate corresponding to (5.29) then takes the form

$$\begin{aligned} |D_{\ell_1} G - \bar{D}_{\ell_1} \bar{G}|_{N-1, K, S+\ell} &\leq |DG(F^{\ell_1} - \bar{F}^{\ell_1})|_{N-1, K, S+\ell} \\ &\quad + |D(G - \bar{G})\bar{F}^{\ell_1}|_{N-1, K, S+\ell} \\ &\leq 2^N |G|_{N-1, K-S-\ell_1, \ell-\ell_1} |F^{\ell_1} - \bar{F}^{\ell_1}|_{N-1, K, S+\ell_1} \\ &\quad + 2^N |G - \bar{G}|_{N-1, K-\ell_1, S+\ell-\ell_1} |\bar{F}^{\ell_1}|_{N-1, K, \ell_1}, \end{aligned}$$

where the estimates in the last inequality are derived just as in (5.29). This reasoning can again be applied iteratively to resolve the entire difference of products of the form $D_{\ell_1} \dots D_{\ell_{i-1}} F^{\ell_i}$, where we note that the loss indices now add up to exactly $S + \ell$ as required. \square

We now turn our attention to the optimally truncated modified vector field $\tilde{F} \equiv \tilde{f}_m \circ \mathbb{P}_m$ where $m = m(h)$ is as in (5.14). This is the same modified vector field which gives rise to the modified Hamiltonian in Lemma 5.10. Then the difference between \tilde{F} and F , the exact vector field of the semilinear evolution equation (1.1), can be estimated as follows.

Lemma 5.15. *Suppose conditions (A), (B0), (B1), and (RK1–3) are satisfied with $K \geq a+b$, where a and b are given by (5.1). Then for every $\delta \in (0, \delta_*]$, the difference between the original vector field F and the modified vectorfield \tilde{F} from Lemma 5.8 satisfies*

$$\|\tilde{F} - F\|_{C_b(\mathcal{D}_{a+b}^{-\delta}; \mathcal{Y})} = O(h^p). \quad (5.33)$$

Proof. We define two intermediate vector fields. Let \tilde{F}^a denote the modified vector field of the numerical method applied to the original semilinear evolution equation (1.1) computed up to order a . Formally, as in the finite dimensional case, it has an expansion of the form (2.4),

$$\tilde{F}^a = F + \sum_{j=p}^{a-1} h^j F^{j+1}. \quad (5.34)$$

Note that $F = \tilde{F}^1$. Due to (5.27) and Lemma 5.11, all coefficients F^k in the expansion (5.34) and consequently \tilde{F}^a are also of class (5.27) with $\ell = a \leq K$.

We now choose h_* and m_* as in the proof of Lemma 5.8 and suppose $m \geq m_*$, so that the projected modified vector fields are well defined. Let $\tilde{F}_m^a \equiv \tilde{f}_m^a \circ \mathbb{P}_m$ denote the corresponding modified vector field of the projected system (5.5), again up to order a , with $m = O(h^{-q/(q+1)})$ as in (5.14). By Lemma 5.11 applied to G_m^ℓ , the vectorfield \tilde{F}_m^a is also of class (5.27).

We now decompose

$$F - \tilde{F} = (F - \tilde{F}^a) + (\tilde{F}^a - \tilde{F}_m^a) + (\tilde{F}_m^a - \tilde{F}). \quad (5.35)$$

We show that when a is chosen as in (5.1), each term on the right is $O(h^p)$ in appropriate norms.

A bound on the first difference on the right of (5.35) follows directly from the definition of \tilde{F}^a in (5.34). Using the norms defined in (5.28),

$$|F - \tilde{F}^a|_{N-1, K, a} \leq h^p \sum_{j=p}^{a-1} h^{j-p} |F^{j+1}|_{N-1, K, a}.$$

By Lemma 5.11, all norms in the right hand sum are finite, so that, for $a \in 1, \dots, K$,

$$|F - \tilde{F}^a|_{N-1, K, a} = O(h^p). \quad (5.36)$$

To estimate the second difference on the right of (5.35), we apply Lemma 5.14 with $\tilde{G}^\ell = G_m^\ell$, which we recall are also of class (5.27), so that (5.32) applies, yielding

$$|\tilde{F}^a - \tilde{F}_m^a|_{N-1, K, a+S} = O(m^{-S}) \quad (5.37)$$

for $S \in 0, \dots, K - a$.

To estimate the third difference on the right of (5.35), as in the proof of Lemma 5.8 we choose $r = \delta/4$ and m_* large enough such that for each $U \in \mathcal{D}_1^{-\delta/2}$, the ball $\mathcal{B}_r^{\mathbb{P}_m \mathcal{Y}_1}(\mathbb{P}_m U)$ is fully contained in \mathcal{D} so that f_m is well-defined on $\mathcal{D} \cap \mathbb{P}_m \mathcal{D}_1$. Then estimate (5.12) holds true and so the bound (2.7) holds for each of the f_m^j with $M = c_F m$ on this ball. We now regard the modified vector field $\tilde{f}_{m(h)}$ with $m(h)$ given by (5.14) as the modified vector field for the differential equation $\dot{u}_m = \tilde{f}_{m(h)}^a(u_m)$ when $\psi_{m(h)}^h$ is used as its $O(h^a)$ time-discretization. In this situation, Lemma 2.2 with a in place of p applies. Moreover, this construction holds true with uniform constants for any $U \in \mathcal{D}_1^{-\delta/2}$, so that, recalling $\tilde{F}_m^a = \tilde{f}_m^a \circ \mathbb{P}_m$ and $\tilde{F} = \tilde{f} \circ \mathbb{P}_m$, the asserted bound reads

$$\|\tilde{F}_{m(h)}^a - \tilde{F}\|_{\mathcal{C}(\mathcal{D}_1^{-\delta/2}; \mathcal{Y}_1)} = O(h^a m(h)^{a+1}). \quad (5.38)$$

We now seek conditions under which the estimates (5.36), (5.37), and (5.38) are all of order h^p . Due to (5.14), $m(h) = O(h^{-q/(1+q)})$, so that the requirement $(m(h))^{-S} = O(h^p)$ leads to the choice

$$S = \left\lceil p \frac{1+q}{q} \right\rceil = b,$$

where b is as in (5.1). Similarly, the requirement $h^a m^{a+1} = O(h^p)$ is equivalent to

$$O(h^a m^{a+1}) = O(h^{a - \frac{q(a+1)}{q+1}}) = O(h^{\frac{a}{q+1} - \frac{q}{q+1}}) = O(h^p)$$

and leads to the choice of a stated in (5.1), namely

$$a = \lceil p(q+1) + q \rceil.$$

Altogether, using (5.35), recalling that $K \geq a + b$, and (5.37) with $S = b$ in (5.37), we obtain (5.33). \square

Remark 5.16. We can change \tilde{F} so that its leading order linear part is A rather than $A\mathbb{P}_{m(h)}$ since, by (3.5), for $U^0 \in \mathcal{Y}_{\tau, \ell}$ the difference between the two modified flows are exponentially small in the \mathcal{Y} -norm.

In the Hamiltonian case, the previous result on $O(h^p)$ -closeness of true and modified vectorfield carries over to a statement on $O(h^p)$ -closeness of the corresponding Hamiltonians.

Lemma 5.17. *Under the assumptions of Lemma 5.15 suppose that, in addition, the semilinear evolution equation is Hamiltonian satisfying (H0–4) and that the numerical method is symplectic (RK4) and of order p . Then the modified Hamiltonian \tilde{H} from Lemma 5.10 can be chosen such that*

$$\|H - \tilde{H}\|_{\mathcal{C}_b(\mathcal{D}_{a+b}^{-\delta}; \mathbb{R})} = O(h^p).$$

Proof. By (H4) and Theorem 3.6, $\mathcal{D}_{a+b}^{-\delta}$ is star-shaped. Let $U^0 \in \text{ck}(\mathcal{D}_{a+b}^{-\delta})$. Since the Hamiltonian from Lemma 5.10 is defined only up to a constant, we can choose the constant of integration such that $H(U_0) = \tilde{H}(U_0)$. Then, following the proof of Lemma 3.5, we estimate, for any $U \in \mathcal{D}_{a+b}^{-\delta}$,

$$\begin{aligned} |\tilde{H}(U) - H(U)| &\leq \left| \int_0^1 \langle \mathbb{J}^{-1}(F - \tilde{F})(tU + (1-t)U_0), U - U_0 \rangle dt \right| \\ &\leq \|\mathbb{J}^{-1}\|_{\mathcal{E}(\mathcal{Y})} \|F - \tilde{F}\|_{\mathcal{C}_b(\mathcal{D}_{a+b}^{-\delta}; \mathcal{Y})} \sup_{U \in \mathcal{D}_{a+b}^{-\delta}} \|U\|_{\mathcal{Y}}. \end{aligned}$$

Since \mathcal{D}_{a+b} is a \mathcal{Y} -bounded set, Lemma 5.15 implies that the right hand side is $O(h^p)$. \square

ACKNOWLEDGMENTS

The authors thank Ernst Hairer, Christian Lubich, and Gerhard Wanner for stimulating discussions and comments, and the Courant Institute of Mathematical Sciences for their hospitality during the initial preparation of this manuscript. C.W. further thanks the University of Geneva for their hospitality and acknowledges funding by the Nuffield Foundation, by the Leverhulme Foundation and by EPSRC grant EP/D063906/1. M.O. was partially supported by a Max-Kade Fellowship, by the ESF network Harmonic and Complex Analysis and Applications (HCAA), and by the German Science Foundation.

REFERENCES

- [1] G. Benettin and A. Giorgilli, *On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms*, J. Statist. Phys. **74** (1994), 1117–1143.
- [2] T. Bridges and S. Reich, *Multi-symplectic integrators: numerical schemes for Hamiltonian PDEs that conserve symplecticity*, Phys. Lett. A **284** (2001), 184–193.
- [3] B. Cano, *Conserved quantities of some Hamiltonian wave equations after full discretization*, Numer. Math. **103** (2006), 197–223.
- [4] D. Cohen, E. Hairer, and C. Lubich, *Conservation of energy, momentum and actions in numerical discretizations of nonlinear wave equations*, Numer. Math. **110** (2008), 113–143.
- [5] A. Debussche and E. Faou, *Modified energy for split-step methods applied to the linear Schrödinger equation*, SIAM J. Numer. Anal. **47** (2009), 3705–3719.
- [6] G. Dujardin and E. Faou, *Normal form and long time analysis of splitting schemes for the linear Schrödinger equation with small potential*, Numer. Math. **106** (2007), 223–262.
- [7] E. Faou and B. Grébert, *Hamiltonian interpolation of splitting approximations for nonlinear PDEs*, Found. Comput. Math. **11** (2011), 381–415.
- [8] E. Faou, B. Grébert, and E. Paturel, *Birkhoff normal form for splitting methods applied to semi linear Hamiltonian PDEs. Part I: Finite dimensional discretization*, Numer. Math. **114** (2010), 429–458.

- [9] E. Faou, B. Grébert, and E. Paturel, *Birkhoff normal form for splitting methods applied to semi linear Hamiltonian PDEs. Part II: Abstract splitting*, Numer. Math. **114** (2010), 459–490.
- [10] E. Faou and B. Grébert, *A Nekhoroshev type theorem for the nonlinear Schrödinger equation on the torus*, Analysis & PDE, to appear.
- [11] A.B. Ferrari and E.S. Titi, *Gevrey regularity for nonlinear analytic parabolic equations*, Commun. Partial Differential Equations **23** (1998), 1–16.
- [12] L. Gauckler and C. Lubich, *Splitting integrators for nonlinear Schrödinger equations over long times*, Found. Comput. Math. **10** (2010), 275–302.
- [13] E. Hairer, *Backward analysis of numerical integrators and symplectic methods*, Annals of Numerical Mathematics **1** (1994), 107–132.
- [14] E. Hairer and C. Lubich, *The life-span of backward error analysis for numerical integrators*, Numer. Math. **76** (1997), 441–462.
- [15] E. Hairer, C. Lubich, and G. Wanner, “Geometric Numerical Integration. Structure-preserving Algorithms for Ordinary Differential Equations,” Springer-Verlag, Berlin, 2002.
- [16] D. Henry, “Geometric Theory of Semilinear Parabolic Equations,” Springer-Verlag, New York, 1983.
- [17] A. Iserles, “First Course in the Numerical Analysis of Differential Equations,” Cambridge University Press, 1996.
- [18] A.L. Islas and C.M. Schober, *Backward error analysis for multisymplectic discretizations of Hamiltonian PDEs*, Math. Comp. Simul. **69** (2005), 290–303.
- [19] B. Leimkuhler and S. Reich, “Simulating Hamiltonian Dynamics,” Cambridge University Press, 2004.
- [20] C. Lubich and A. Ostermann, *Runge–Kutta methods for parabolic equations and convolution quadrature*, Math. Comp. **60** (1993), 105–131.
- [21] J.E. Marsden and T.S. Ratiu, “Introduction to Mechanics and Symmetry,” Springer-Verlag, New York, 1994.
- [22] K. Matthies, *Time-averaging under fast periodic forcing of parabolic partial differential equations: Exponential estimates*, J. Diff. Eqns. **174** (2001), 88–133.
- [23] K. Matthies and A. Scheel, *Exponential averaging of Hamiltonian evolution equations*, Trans. Amer. Math. Soc. **355** (2003), 747–773.
- [24] B. Moore and S. Reich, *Backward error analysis for Hamiltonian PDEs with applications to nonlinear wave equations*, Numer. Math. **95** (2003), 625–652.
- [25] A.I. Neishtadt, *On the separation of motions in systems with rapidly rotating phase*, J. Appl. Math. Mech. **48** (1984), 134–139.
- [26] M. Oliver, M. West, and C. Wulff, *Approximate momentum conservation for spatial semidiscretizations of nonlinear wave equations*, Numer. Math. **97** (2004), 493–535.
- [27] M. Oliver and C. Wulff, *A-stable Runge–Kutta methods for semilinear evolution equations on scales of Banach spaces*, J. Functional Anal. **263** (2012), 1981–2023.
- [28] M. Oliver and C. Wulff, *Stability under Galerkin truncation of A-stable Runge–Kutta discretizations in time* submitted for publication (2011), arXiv:1007.4712 [math.NA].
- [29] A. Pazy, “Semigroups of Linear Operators and Applications to Partial Differential Equations,” Springer-Verlag, New York, 1983.
- [30] S. Kuksin and J. Pöschel, *Invariant Cantor Manifolds of quasi-periodic oscillations for nonlinear Schrödinger equation*, Ann. Math. **143** (1996) 149–179.
- [31] J. Pöschel, *Quasi-periodic solutions for nonlinear wave equations*, Comm. Math. Helv. **71** (1996), 269–296.
- [32] M. Reed and B. Simon, “Methods of Modern Mathematical Physics. I. Functional Analysis,” Academic Press, San Diego, 1972.
- [33] S. Reich, *Backward error analysis for numerical integrators*, SIAM J. Numer. Anal. **36** (1999), 1549–1570.
- [34] C.R. Smith, *A characterization of star-shaped sets*, Am. Math. Mon. **75** (1968), 386.

(C. Wulff) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SURREY, GUILDFORD GU2 7XH,
UK

E-mail address: `c.wulff@surrey.ac.uk`

(M. Oliver) SCHOOL OF ENGINEERING AND SCIENCE, JACOBS UNIVERSITY, 28759 BREMEN,
GERMANY

E-mail address: `oliver@member.ams.org`